PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/12, C07K 14/78, 14/47, C12N 15/63, A01K 67/027, C07K 16/18, C12Q 1/68, A61K 38/17, G01N 33/68

A2

(11) International Publication Number: WO 00/68380

(43) International Publication Date: 16 November 2000 (16.11.00)

(21) International Application Number: PCT/US00/12811

(22) International Filing Date: 10 May 2000 (10.05.00)

(30) Priority Data:
60/133,643 11 May 1999 (11.05.99) US
60/150,409 23 August 1999 (23.08.99) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications
US 60/133,643 (CIP)
Filed on 11 May 1999 (11.05.99)
US 60/150,409 (CIP)
Filed on 23 August 1999 (23.08.99)

(71) Applicant (*for all designated States except US*): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #12, Mountain View, CA 94040 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94086 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: EXTRACELLULAR MATRIX AND ADHESION-ASSOCIATED PROTEINS

(57) Abstract

The invention provides human extracellular matrix and adhesion-associated proteins (EXMAD) and polynucleotides which identify and encode EXMAD. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of EXMAD.

B8

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PC

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

EXTRACELLULAR MATRIX AND ADHESION-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of extracellular matrix and adhesion-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, immune, reproductive, neuronal, and genetic disorders.

BACKGROUND OF THE INVENTION

Extracellular Matrix Proteins

The extracellular matrix (ECM) is a complex network of glycoproteins, polysaccharides, proteoglycans, and other macromolecules that are secreted from the cell into the extracellular space. The ECM remains in close association with the cell surface and provides a supportive meshwork that profoundly influences cell shape, motility, strength, flexibility, and adhesion. In fact, adhesion of a cell to its surrounding matrix is required for cell survival except in the case of metastatic tumor cells, which have overcome the need for cell-ECM anchorage. This phenomenon suggests that the ECM plays a critical role in the molecular mechanisms of growth control and metastasis. (Reviewed in Ruoslahti, E. (1996) Sci. Am. 275:72-77.) Furthermore, the ECM determines the structure and physical properties of connective tissue and is particularly important for morphogenesis and other processes associated with embryonic development and pattern formation.

Collagens

The collagens comprise a family of ECM proteins that provide structure to bone, teeth, skin, ligaments, tendons, cartilage, blood vessels, and basement membranes. Multiple collagen proteins have been identified. Three collagen molecules fold together in a triple helix stabilized by interchain disulfide bonds. Bundles of these triple helices then associate to form fibrils. Collagen primary structure consists of hundreds of (Gly-X-Y) repeats where about a third of the X and Y residues are Pro. Glycines are crucial to helix formation as the bulkier amino acid side chains cannot fold into the triple helical conformation. Because of these strict sequence requirements, mutations in collagen genes have severe consequences. Osteogenesis imperfecta patients have brittle bones that fracture easily; in severe cases patients die in utero or at birth. Ehler-Danlos syndrome patients have hyperelastic skin, hypermobile joints, and susceptibility to aortic and intestinal rupture. Chondrodysplasia patients have short stature and ocular disorders. Alport syndrome patients have hematuria, sensorineural deafness, and eye lens deformation. (See Isselbacher, K.J., et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc., New York, NY, pp. 2105-2117; and Creighton, T.E. (1984) Proteins,

Structures and Molecular Principles, W.H. Freeman and Company, New York, NY, pp. 191-197.),

Collectins are extracellular proteins with collagen tails and globular lectin domains that play an important role in the first line immune response to micoorganisms. The peripheral lectin domain permits binding to sugar residues on microorganisms, while the collagen tail interacts with phagocyte receptors or the complement system. Examples of collectins are the pulmonary surfactant proteins SP-A and SP-D (Kuroki, S.D. et al. (1998) J. Biol. Chem. 273:4783-4789).

Elastin

Elastin and related proteins confer elasticity to tissues such as skin, blood vessels, and lungs. Elastin is a highly hydrophobic protein of about 750 amino acids that is rich in proline and glycine residues. Elastin molecules are highly cross-linked, forming an extensive extracellular network of fibers and sheets. Elastin fibers are surrounded by a sheath of microfibrils which are composed of a number of glycoproteins, including fibrillin. Mutations in the gene encoding fibrillin are responsible for Marfan's syndrome, a genetic disorder characterized by defects in connective tissue. In severe cases, the aortas of afflicted individuals are prone to rupture. (Reviewed in Alberts, B., et al. (1994) Molecular Biology of the Cell, Garland Publishing, New York, NY, pp. 984-986.)

Fibronectin

Fibronectin is a large ECM glycoprotein found in all vertebrates. Fibronectin exists as a dimer of two subunits, each containing about 2,500 amino acids. Each subunit folds into a rod-like structure containing multiple domains. The domains each contain multiple repeated modules, the most common of which is the type III fibronectin repeat. The type III fibronectin repeat is about 90 amino acids in length and is also found in other ECM proteins and in some plasma membrane and cytoplasmic proteins. Furthermore, some type III fibronectin repeats contain a characteristic tripeptide consisting of Arginine-Glycine-Aspartic acid (RGD). The RGD sequence is recognized by the integrin family of cell surface receptors and is also found in other ECM proteins. Disruption of both copies of the gene encoding fibronectin causes early embryonic lethality in mice. The mutant embryos display extensive morphological defects, including defects in the formation of the notochord, somites, heart, blood vessels, neural tube, and extraembryonic structures. (Reviewed in Alberts, supra, pp. 986-987.)

Laminin

Laminin is a major glycoprotein component of the basal lamina which underlies and supports epithelial cell sheets. Laminin is one of the first ECM proteins synthesized in the developing embryo. Laminin is an 850 kilodalton protein composed of three polypeptide chains joined in the shape of a

cross by disulfide bonds. Laminin is especially important for angiogenesis and, in particular, for guiding the formation of capillaries. (Reviewed in Alberts, supra, pp. 990-991.)

Proteoglycans

There are many other types of proteinaceous ECM components, most of which can be classified as proteoglycans. Proteoglycans are composed of unbranched polysaccharide chains (glycosaminoglycans) attached to protein cores. Common proteoglycans include aggrecan, betaglycan, decorin, perlecan, serglycin, and syndecan-1. Some of these molecules not only provide mechanical support, but also bind to extracellular signaling molecules, such as fibroblast growth factor and transforming growth factor β, suggesting a role for proteoglycans in cell-cell communication. (Reviewed in Alberts, supra, pp. 973-978.) Likewise, the glycoproteins tenascin-C and tenascin-R are expressed in developing and lesioned neural tissue and provide stimulatory and anti-adhesive (inhibitory) properties, respectively, for axonal growth (Faissner, A. (1997) Cell Tissue Res. 290:331-341).

Dentin phosphoryn (DPP) is a major component of the dentin ECM. DPP is a proteoglycan that is synthesized and expressed by odontoblasts (Gu, K., et al. (1998) Eur. J. Oral Sci. 106:1043-1047). DPP is believed to nucleate or modulate the formation of hydroxyapatite crystals. The gene encoding DPP has been mapped to human chromosome 4. Chromosome 4 contains the gene loci for two dentin genetic diseases, dentinogenesis imperfecta type II and dentin dysplasia type II (Feng, J.Q., et al. (1998) J. Biol. Chem. 273:9457-9464).

Mucins

Mucins are highly glycosylated glycoproteins that are the major structural component of the mucus gel. The physiological functions of mucins are cytoprotection, mechanical protection, maintenance of viscosity in secretions, and cellular recognition. MUC6 is a human gastric mucin that is also found in gall bladder, pancreas, seminal vesicles, and female reproductive tract (Toribara, N.W., et al. (1997) J. Biol. Chem. 272:16398-16403). The MUC6 gene has been mapped to human chromosome 11 (Toribara, N.W., et al. (1993) J. Biol. Chem. 268:5879-5885). Hemomucin is a novel Drosophila surface mucin that may be involved in the induction of antibacterial effector molecules (Theopold, U., et al. (1996) J. Biol. Chem. 217:12708-12715).

Link Protein

Link protein binds to both cartilage proteoglycan and hyaluronan in cartilage ECM. This binding stabilizes the aggregation of these cartilage ECM proteins and produces supramolecular

assemblies. Link protein has been detected in other connective tissues, where it may bind proteoglycans and hyaluronan. Link protein contains a signal peptide, an immunoglobulin repeat, and link repeats (Ayad, S., et al. (1994) The Extracellular Matrix Facts Book, Academic Press, Inc., San Diego, CA, pp. 120-121).

Adhesion-Associated Proteins

The surface of a cell is rich in transmembrane proteoglycans, glycoproteins, glycolipids, and receptors. These macromolecules mediate adhesion with other cells and with components of the ECM. The interaction of the cell with its surroundings profoundly influences cell shape, strength, flexibility, motility, and adhesion. These dynamic properties are intimately associated with signal transduction pathways controlling cell proliferation and differentiation, tissue construction, and embryonic development.

Cadherins

Cadherins comprise a family of calcium-dependent glycoproteins that function in mediating cell-cell adhesion in virtually all solid tissues of multicellular organisms. These proteins share multiple repeats of a cadherin-specific motif, and the repeats form the folding units of the cadherin ECM. Cadherin molecules cooperate to form focal contacts, or adhesion plaques, between adjacent epithelial cells. The cadherin family includes the classical cadherins and protocadherins. Classical cadherins include the E-cadherin, N-cadherin, and P-cadherin subfamilies. E-cadherin is present on many types of epithelial cells and is especially important for embryonic development. P-cadherin is present on cells of the placenta and epidermis. Recent studies report that protocadherins are involved in a variety of cell-cell interactions (Suzuki, S. T. (1996) J. Cell Sci. 109:2609-2611). The intracellular anchorage of cadherins is regulated by their dynamic association with catenins, a family of cytoplasmic signal transduction proteins associated with the actin cytoskeleton. The anchorage of cadherins to the actin cytoskeleton appears to be regulated by protein tyrosine phosphorylation, and the cadherins are the target of phosphorylation-induced junctional disassembly (Aberle, H., et al. (1996) J. Cell. Biochem. 61:514-523).

Integrins

Integrins are ubiquitous transmembrane adhesion molecules that link the ECM to the internal cytoskeleton. Integrins are composed of two noncovalently associated transmembrane glycoprotein subunits called α and β. Integrins function as receptors that play a role in signal transduction. For example, binding of integrin to its extracellular ligand may stimulate changes in intracellular calcium

levels or protein kinase activity (Sjaastad, M.D. and Nelson, W.J. (1997) BioEssays 19:47-55). At least ten cell surface receptors of the integrin family recognize the ECM component fibronectin, which is involved in many different biological processes including cell migration and embryogenesis (Johansson, S., et al. (1997) Front. Biosci. 2:D126-D146).

Lectins

Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells. (Reviewed in Drickamer, K. and Taylor, M.E. (1993) Annu. Rev. Cell Biol. 9:237-264.) This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L.A. (1991) J. Cell. Biochem. 45:139-146; Paietta, E., et al. (1989) J. Immunol. 143:2850-2857).

Lectins are further classified into subfamilies based on carbohydrate-binding specificity and other criteria. The galectin subfamily, in particular, includes lectins that bind β-galactoside carbohydrate moieties in a thiol-dependent manner. (Reviewed in Hadari, Y.R., et al. (1998) J. Biol. Chem. 270:3447-3453.) Galectins are widely expressed and developmentally regulated. Because all galectins lack an N-terminal signal peptide, it is suggested that galectins are externalized through an atypical secretory mechanism. Two classes of galectins have been defined based on molecular weight and oligomerization properties. Small galectins form homodimers and are about 14-16 kilodaltons in mass, while large galectins are monomeric and about 29-37 kilodaltons.

Galectins contain a characteristic carbohydrate recogntion domain (CRD). The CRD is about 140 amino acids and contains several stretches of about 1-10 amino acids which are highly conserved among all galectins. A particular 6-amino acid motif within the CRD contains conserved tryptophan and arginine residues which are critical for carbohydrate binding. The CRD of some galectins also contains cysteine residues which may be important for disulfide bond formation. Secondary structure predictions indicate that the CRD forms several β-sheets.

Galectins play a number of roles in diseases and conditions associated with cell-cell and cell-matrix interactions. For example, certain galectins associate with sites of inflammation and bind to cell surface immunoglobulin E molecules. In addition, galectins may play an important role in cancer metastasis. Galectin overexpression is correlated with the metastatic potential of cancers in humans and mice. Moreover, anti-galectin antibodies inhibit processes associated with cell transformation, such as cell aggregation and anchorage-independent growth. (See, for example, Su, Z.-Z., et al. (1996) Proc. Natl. Acad. Sci. USA 93:7252-7257.)

Selectins

Selectins, or LEC-CAMs, comprise a specialized lectin subfamily involved primarily in inflammation and leukocyte adhesion. (Reviewed in Lasky; supra.) Selectins, which mediate the recruitment of leukocytes from the circulation to sites of acute inflammation, are expressed on the surface of vascular endothelial cells in response to cytokine signaling. Selectins bind to specific ligands on the leukocyte cell membrane and enable the leukocyte to adhere to and migrate along the endothelial surface. Binding of selectin to its ligand leads to polarized rearrangement of the actin cytoskeleton and stimulates signal transduction within the leukocyte (Brenner, B., et al. (1997) Biochem. Biophys. Res. Commun. 231:802-807; Hidari, K.I., et al. (1997) J. Biol. Chem. 272:28750-28756). Members of the selectin family possess three characteristic motifs: a lectin or carbohydrate recognition domain; an epidermal growth factor (EGF)-like domain; and a variable number of short consensus repeats (scr or "sushi" repeats) which are also present in complement regulatory proteins. The selectins include lymphocyte adhesion molecule-1 (LAM-1 or L-selectin), endothelial leukocyte adhesion molecule-1 (ELAM-1 or E-selectin), and granule membrane protein-140 (GMP-140 or P-selectin) (Johnston, G.I., et al. (1989) Cell 56:1033-1044).

Attractin

Attractin is a 134 kilodalton glycoprotein found in the serum. It is a member of the CUB family of cell adhesion proteins and binds directly to leukocytes. Attractin has a CUB domain, an EGF domain, and C-type lectin protein domains. This serum protein mediates the interaction between T lymphocytes and monocytes and leads to the adherence and spreading of monocytes that become the foci for T cell clustering. (See, Duke-Cohan, J.S., et al. (1998) Proc. Natl. Acad. Sci. USA 95:11336-11341.)

Proteins Containing Leucine Rich Repeats (LRRs)

LRRs are sequence motifs, approximately 22-28 amino acids in length, found in proteins with a large variety of functions and cellular locations. Proteins containing LRRs are all thought to be involved in protein-protein interactions. The crystal structure of LRRs has been studied and found to correspond to beta-alpha structural units. These structural units form a parallel beta sheet with one surface exposed to solvent. In this way an LRR-containing protein acquires a nonglobular shape (Kobe, B. and Deisenhofer, J. (1994) Trends Biochem. Sci. 19:415-421). There is evidence to suggest LRRs function in signal transduction and cellular adhesion as well as in protein-protein interactions (Gay, N.J., et al. (1991) FEBS Lett. 29:87-91). For example, LLR proteins such as connectin and chaoptin are important cell adhesion molecules in neuronal development in Drosophilia melanogaster.

and mammalian homologs are found in mouse (Taguchi, et al. (1996) Brain Res.Mol. Brain Res. 1-2:31-40).

Proteins Containing Armadillo/β-Catenin-like Repeats

Various proteins such as those encoded by the Drosophila armadillo gene and the human APC gene contain amino acid repeats that interact with β-catenins. The armadillo gene is required for pattern formation within the embryonic segments and imaginal discs and is highly conserved. It is 63% identical to a human protein, plakoglobin, which is involved in adhesive junctions joining epithelial and other cells (Peifer, M. and Wieschaus, E. (1990) Cell 63:1167-1176). APC gene mutations appear to initiate inherited forms of human colorectal cancer and sporadic forms of colorectal and gastric cancer (Rubinfeld, B., et al. (1993) Science 262:1731-1734). The fact that the protein encoded by APC interacts with catenin suggests a link between tumor initiation and cell adhesion (Su, L.K., et al. (1993) Science 262:1734-1737).

Proteins Containing C-type Lectin Domains

C-type lectin domains are found in a variety of proteins, including selectins and lecticans. Lecticans are a family of chondroitin sulfate proteoglycans that include aggrecan, versican, neurocan, and brevican. All C-type lectin proteins are involved in protein-protein interactions (Aspberg, A., et al. (1997) Proc. Natl. Acad. Sci. USA 94:10116-10121). A novel macrophage-restricted C-type lectin protein has been cloned from mouse tissue. It is a type II transmembrane protein with one extracellular C-type lectin domain (Balch, S.G., et al. (1998) J. Biol. Chem. 273:18656-18664).

Bystin

Bystin is a cytoplasmic protein that binds directly to trophinin, a cell adhesion molecule, and tastin. The three molecules form a complex that is involved in cell adhesion. Bystin, tastin, and trophinin are strongly expressed in cells involved in the implantation of embryos, specifically in cells at human implantation sites and in intermediate trophoblasts at the invasion front of the placenta in early pregnancy. Bystin also binds to cytokeratins. During early embryogenesis cytokeratins 8 and 18 are expressed in the trophectoderm of blastocytes. It is possible that the molecular complex formed by bystin, tastin, and trophinin interacts with the cytokeratins of trophectoderm cells at the time of implantation. A key component of embryo implantation is the unique cell adhesion to endometrial epithelium that occurs and the subsequent invasion of the maternal tissue by the trophoblast. Bystin may have an important role in the signal transduction that links cell adhesion to proliferation (Suzuki, N., et al. (1998) Proc. Natl. Acad. Sci. 95:5027-5032).

Src-homology 3 (SH3) Domain-Containing Proteins

SH3 is a 60-70 amino acid motif found in a variety of signal transduction and cytoskeletal proteins. The SH3 domain is involved in mediating protein-protein interactions. Evidence suggests that the SH3 domains recognize a family of related domains or proteins in a variety of different tissues and species. One novel SH3 domain-containing protein is the 52 kilodalton focal adhesion protein (FAP52 or p52). FAP52 is localized to focal adhesions, specialized membrane domains in cultured cells that mediate the attachment of cells to the growth substratum and ECM. Focal adhesions consist of structural proteins, integrins, regulatory molecules, and signaling molecules and are involved in cell signaling. FAP52 may form part of this multimolecular complex that comprises focal adhesion sites (Merilainent, J., et al. (1997) J. Biol. Chem. 272:23278-23284).

The ECM plays an important role in cell invasive processes such as angiogenesis and tumor metastasis (Ruoslahti, supra). In particular, the glycoproteins laminin and fibronectin are implicated in the migration of tumor cells through the ECM (chemotaxis) in the course of metastasis of tumors to other tissues. The same process, chemotaxis, also promotes the migration of vascular endothelial cells to form new microvascular networks to support these tumors (tumor angiogenesis).

The discovery of new extracellular matrix and adhesion-associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, immune, reproductive, neuronal, and genetic disorders.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, extracellular matrix and adhesion-associated proteins, referred to collectively as "EXMAD" and individually as "EXMAD-1," "EXMAD-2," "EXMAD-3," "EXMAD-4," "EXMAD-5," "EXMAD-6," "EXMAD-7,", "EXMAD-8," "EXMAD-9," "EXMAD-10," "EXMAD-11," "EXMAD-12," "EXMAD-13," "EXMAD-14," "EXMAD-15," "EXMAD-16," "EXMAD-17," "EXMAD-18," "EXMAD-19," "EXMAD-20," "EXMAD-21," "EXMAD-22," "EXMAD-23," "EXMAD-24," and "EXMAD-25." In one aspect, the invention provides an isolated polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-25.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. In one alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:26-50.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

The invention further provides an isolated polynucleotide comprising a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, b) a naturally occurring

polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, c) a polynucleotide sequence complementary to a), or d) a polynucleotide sequence complementary to b). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, c) a polynucleotide sequence complementary to a), or d) a polynucleotide sequence complementary to b). The method comprises a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 30 contiguous nucleotides. In another alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and a pharmaceutically acceptable excipient. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional EXMAD, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a

pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional EXMAD, comprising administering to a patient in need of such treatment the pharmaceutical composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional EXMAD, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:26-50, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding EXMAD.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of EXMAD.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding EXMAD were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze EXMAD, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"EXMAD" refers to the amino acid sequences of substantially purified EXMAD obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of EXMAD. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of EXMAD either by directly interacting with EXMAD or by acting on components of the biological pathway in which EXMAD participates.

An "allelic variant" is an alternative form of the gene encoding EXMAD. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times i a given sequence.

"Altered" nucleic acid sequences encoding EXMAD include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as EXMAD or a polypeptide with at least one functional characteristic of EXMAD. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding EXMAD, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding EXMAD. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent EXMAD. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of EXMAD is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of EXMAD. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of EXMAD either by directly interacting with EXMAD or by acting on components of the biological pathway in which EXMAD participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind EXMAD polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired.

Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic EXMAD, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or

amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding EXMAD or fragments of EXMAD may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the

side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "fragment" is a unique portion of EXMAD or the polynucleotide encoding EXMAD which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:26-50 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:26-50, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:26-50 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:26-50 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:26-50 and the region of SEQ ID NO:26-50 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-25 is encoded by a fragment of SEQ ID NO:26-50. A fragment of SEQ ID NO:1-25 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-25. For example, a fragment of SEQ ID NO:1-25 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-25. The precise length of a fragment of SEQ ID NO:1-25 and the region of SEQ ID NO:1-25 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended

purpose for the fragment.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may subsitute for the word "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis

programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameter

of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding

between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune

disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of EXMAD which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of EXMAD which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of EXMAD. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of EXMAD.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding EXMAD, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid

sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers ma be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as prime binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides fc microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary

polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding EXMAD, or fragments thereof, or EXMAD itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, and plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may

possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human extracellular matrix and adhesion-associated proteins (EXMAD), the polynucleotides encoding EXMAD, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, immune, reproductive, neuronal, and genetic disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding EXMAD. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each EXMAD were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each EXMAD and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions

associated with nucleotide sequences encoding EXMAD. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:26-50 and to distinguish between SEQ ID NO:26-50 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express EXMAD as a fraction of total tissues expressing EXMAD. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing EXMAD as a fraction of total tissues expressing EXMAD. Column 5 lists the vectors used to subclone each cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding EXMAD were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:42 maps to chromosome 8 within the interval from 64.60 to 90.20 centiMorgans. SEQ ID NO:48 maps to chromosome 2 within the interval from 193.60 to 197.60 centiMorgans.

The invention also encompasses EXMAD variants. A preferred EXMAD variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the EXMAD amino acid sequence, and which contains at least one functional or structural characteristic of EXMAD.

The invention also encompasses polynucleotides which encode EXMAD. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:26-50, which encodes EXMAD. The polynucleotide sequences of SEQ ID NO:26-50, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding EXMAD. In particular, such a variant polynucleotide sequence will have at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding EXMAD. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:26-50 which has at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:26-50. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of EXMAD.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding EXMAD, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring EXMAD, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode EXMAD and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring EXMAD under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding EXMAD or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding EXMAD and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode EXMAD and EXMAD derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding EXMAD or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:26-50 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with

machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding EXMAD may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode EXMAD may be cloned in recombinant DNA molecules that direct expression of EXMAD, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express EXMAD.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter EXMAD-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of EXMAD, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby

maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding EXMAD may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, EXMAD itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of EXMAD, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active EXMAD, the nucleotide sequences encoding EXMAD or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding EXMAD. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding EXMAD. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding EXMAD and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding EXMAD and appropriate transcriptional and translational

control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding EXMAD. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding EXMAD. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding EXMAD can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding EXMAD into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of EXMAD are needed, e.g. for the production of antibodies, vectors which direct high level expression of EXMAD may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of EXMAD. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of EXMAD. Transcription of sequences encoding EXMAD may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J.

6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding EXMAD may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses EXMAD in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of EXMAD in cell lines is preferred. For example, sequences encoding EXMAD can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980)

Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding EXMAD is inserted within a marker gene sequence, transformed cells containing sequences encoding EXMAD can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding EXMAD under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding EXMAD and that express EXMAD may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of EXMAD using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on EXMAD is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding EXMAD include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding EXMAD, or any fragments thereof, may be cloned into a vector

for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding EXMAD may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode EXMAD may be designed to contain signal sequences which direct secretion of EXMAD through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding EXMAD may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric EXMAD protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of EXMAD activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize

these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the EXMAD encoding sequence and the heterologous protein sequence, so that EXMAD may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled EXMAD may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

Fragments of EXMAD may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of EXMAD may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of EXMAD and extracellular matrix and adhesion-associated proteins. In addition, the expression of EXMAD is closely associated with cancerous, proliferating, inflamed, nervous, reproductive, urologic, hematopoietic/immune, cardiovascular, musculoskeletal, developmental, and gastrointestinal tissues, and with cell proliferative disorders, including cancer, inflammation and the immune response. Therefore, EXMAD appears to play a role in cell proliferative, immune, reproductive, neuronal, and genetic disorders. In the treatment of disorders associated with increased EXMAD expression or activity, it is desirable to decrease the expression or activity of EXMAD. In the treatment of disorders associated with decreased EXMAD expression or activity, it is desirable to increase the expression or activity of EXMAD.

Therefore, in one embodiment, EXMAD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of EXMAD. Examples of such disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate,

salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune disorder, such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a reproductive disorder, such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neuronal disorder, such as akathesia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, diabetic neuropathy, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, peripheral neuropathy, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, postherpetic neuralgia, schizophrenia, and Tourette's disorder; and a genetic disorder, such as adrenoleukodystrophy, Alport's syndrome, choroideremia, Duchenne and Becker muscular dystrophy, Down's syndrome, cystic fibrosis, chronic granulomatous disease, dentinogenesis imperfecta type II, dentin dysplasia type II, Gaucher's disease, Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic dystrophy, pycnodysostosis, Refsum's syndrome retinoblastoma, sickle cell anemia, thalassemia, Werner syndrome, von Willebrand's disease, Wilms tumor, Zellweger syndrome, peroxisomal acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, peroxisomal bifunctional protein deficiency, mitochondrial carnitine palmitoyl transferase

and carnitine deficiency, mitochondrial very-long-chain acyl-CoA dehydrogenase deficiency, mitochondrial medium-chain acyl-CoA dehydrogenase deficiency, mitochondrial short-chain acyl-CoA dehydrogenase deficiency, mitochondrial electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency.

In another embodiment, a vector capable of expressing EXMAD or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of EXMAD including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EXMAD in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of EXMAD including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EXMAD may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of EXMAD including, but not limited to, those listed above.

In a further embodiment, an antagonist of EXMAD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of EXMAD. Examples of such disorders include, but are not limited to, those cell proliferative, immune, reproductive, neuronal, and genetic disorders described above. In one aspect, an antibody which specifically binds EXMAD may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express EXMAD.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding EXMAD may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of EXMAD including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of EXMAD may be produced using methods which are generally known in the art. In particular, purified EXMAD may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind EXMAD. Antibodies to EXMAD may also be generated using methods that are well known in the art. Such antibodies may include, but are

not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with EXMAD or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to EXMAD have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of EXMAD amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to EXMAD may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce EXMAD-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in

the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for EXMAD may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between EXMAD and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering EXMAD epitopes is generally used, but a competitive binding assay may also be employed (Pound, <u>supra</u>).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for EXMAD. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of EXMAD-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple EXMAD epitopes, represents the average affinity, or avidity, of the antibodies for EXMAD. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular EXMAD epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the EXMAD-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of EXMAD, preferably in active form, from the antibody (Catty, D. (1988) <u>Antibodies, Volume I: A Practical Approach</u>, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) <u>A Practical Guide to Monoclonal Antibodies</u>, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of EXMAD-antibody

complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding EXMAD, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding EXMAD may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding EXMAD. Thus, complementary molecules or fragments may be used to modulate EXMAD activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding EXMAD.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding EXMAD. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding EXMAD can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding EXMAD. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding EXMAD. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding EXMAD.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by <u>in vitro</u> and <u>in vivo</u> transcription of DNA sequences encoding EXMAD. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use <u>in vivo</u>, <u>in vitro</u>, and <u>ex vivo</u>. For <u>ex vivo</u> therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such

therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of EXMAD, antibodies to EXMAD, and mimetics, agonists, antagonists, or inhibitors of EXMAD. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of <u>Remington's Pharmaceutical Sciences</u> (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene

glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of EXMAD, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the

active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example EXMAD or fragments thereof, antibodies of EXMAD, and agonists, antagonists or inhibitors of EXMAD, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind EXMAD may be used for the

diagnosis of disorders characterized by expression of EXMAD, or in assays to monitor patients being treated with EXMAD or agonists, antagonists, or inhibitors of EXMAD. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for EXMAD include methods which utilize the antibody and a label to detect EXMAD in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring EXMAD, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of EXMAD expression. Normal or standard values for EXMAD expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to EXMAD under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of EXMAD expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding EXMAD may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of EXMAD may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of EXMAD, and to monitor regulation of EXMAD levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding EXMAD or closely related molecules may be used to identify nucleic acid sequences which encode EXMAD. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding EXMAD, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the EXMAD encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:26-50 or from genomic sequences including promoters, enhancers, and introns of the EXMAD gene.

Means for producing specific hybridization probes for DNAs encoding EXMAD include the

cloning of polynucleotide sequences encoding EXMAD or EXMAD derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding EXMAD may be used for the diagnosis of disorders associated with expression of EXMAD. Examples of such disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an immune disorder, such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; a reproductive disorder, such as a disorder of prolacti production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, a

galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neuronal disorder, such as akathesia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, diabetic neuropathy, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, peripheral neuropathy, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, postherpetic neuralgia, schizophrenia, and Tourette's disorder; and a genetic disorder, such as adrenoleukodystrophy, Alport's syndrome, choroideremia, Duchenne and Becker muscular dystrophy, Down's syndrome, cystic fibrosis, chronic granulomatous disease, dentinogenesis imperfecta type II, dentin dysplasia type II, Gaucher's disease, Huntington's chorea, Marfan's syndrome, muscular dystrophy, myotonic dystrophy, pycnodysostosis, Refsum's syndrome, retinoblastoma, sickle cell anemia, thalassemia, Werner syndrome, von Willebrand's disease, Wilms' tumor, Zellweger syndrome, peroxisomal acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, peroxisomal bifunctional protein deficiency, mitochondrial carnitine palmitoyl transferase and carnitine deficiency, mitochondrial very-long-chain acyl-CoA dehydrogenase deficiency, mitochondrial medium-chain acyl-CoA dehydrogenase deficiency, mitochondrial short-chain acyl-CoA dehydrogenase deficiency, mitochondrial electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, mitochondrial trifunctional protein deficiency, and mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency. The polynucleotide sequences encoding EXMAD may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered EXMAD expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding EXMAD may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding EXMAD may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding EXMAD in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of EXMAD, a normal or standard profile for expression is established. This may be accomplished by

combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding EXMAD, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from norma subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding EXMAD may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding EXMAD, or a fragment of a polynucleotide complementary to the polynucleotide encoding EXMAD, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of EXMAD include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presente in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray ca be used to monitor the expression level of large numbers of genes simultaneously and to identify genet

variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding EXMAD may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding EXMAD on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or

affected individuals.

In another embodiment of the invention, EXMAD, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between EXMAD and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with EXMAD, or fragments thereof, and washed. Bound EXMAD is then detected by methods well known in the art. Purified EXMAD can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding EXMAD specifically compete with a test compound for binding EXMAD. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with EXMAD.

In additional embodiments, the nucleotide sequences which encode EXMAD may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No.60/133,643 and U.S. Ser. No.60/150,409 are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic

solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-

well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation

using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:26-50. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding EXMAD occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous,

reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of EXMAD Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:40-50 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:40-50 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:42 and SEQ ID NO:48 are described in The Invention as ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters.

VI. Extension of EXMAD Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:26-50 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR

was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE

Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:26-50 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:26-50 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or

fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

IX. Complementary Polynucleotides

Sequences complementary to the EXMAD-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring EXMAD. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of EXMAD. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the EXMAD-encoding transcript.

X. Expression of EXMAD

Expression and purification of EXMAD is achieved using bacterial or virus-based expression systems. For expression of EXMAD in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express EXMAD upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of EXMAD in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding EXMAD by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, EXMAD is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from EXMAD at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified EXMAD obtained by these methods can be used directly in the following activity assay.

XI. Demonstration of EXMAD Activity

An assay for EXMAD activity measures the disruption of cytoskeletal filament networks upon overexpression of EXMAD in cultured cell lines. (Rezniczek, G. A. et al. (1998) J. Cell Biol. 141:209-225.) cDNA encoding EXMAD is subcloned into a mammalian expression vector that drives high levels of cDNA expression. This construct is transfected into cultured cells, such as rat kangaroo PtK2 or rat bladder carcinoma 804G cells. Actin filaments and intermediate filaments such as keratin and vimentin are visualized by immunofluorescence microscopy using antibodies and techniques well known in the art. The configuration and abundance of cyoskeletal filaments can be assessed and quantified using confocal imaging techniques. In particular, the bundling and collapse of cytoskeletal filament networks are indicative of EXMAD activity.

Alternatively, an assay for EXMAD activity measures the amount of cell aggregation induced by overexpression of EXMAD. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding EXMAD contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (Clontech), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that associated with untransfected cells. The amount of cell agglutination is a direct measure of EXMAD activity.

Alternatively, cell adhesion activity in EXMAD is measured in a 96-well plate assay in which wells are first coated with EXMAD by adding solutions of EXMAD of varying concentrations to the wells. Excess EXMAD is washed off with saline, and the wells incubated with a solution of 1% bovine serum albumin to block non-specific cell binding. Aliquots of a cell suspension of a suitable cell type are then added to the wells and incubated for a period of time at 37 °C. Non-adhered cells are washed

off with saline and the cells stained with a suitable cell stain such as Coomassie blue. The intensity of staining is measured using a variable wavelength 96-well plate reader and compared to a standard curve to determine the number of cells adhering to the EXMAD coated plates. The degree of cell staining is proportional to the cell adhesion activity of EXMAD in the sample.

Alternatively, EXMAD activity is also measured by the interaction of EXMAD with other molecules. EXMAD, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) Biochem. J. 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled EXMAD, washed, and any wells with labeled EXMAD complex are assayed. Data obtained using different concentrations of EXMAD are used to calculate values for the number, affinity, and association of EXMAD with the candidate molecules.

XII. Functional Assays

EXMAD function is assessed by expressing the sequences encoding EXMAD at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of EXMAD on gene expression can be assessed using highly purified

populations of cells transfected with sequences encoding EXMAD and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding EXMAD and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of EXMAD Specific Antibodies

EXMAD substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the EXMAD amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-EXMAD activity by, for example, binding the peptide or EXMAD to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring EXMAD Using Specific Antibodies

Naturally occurring or recombinant EXMAD is substantially purified by immunoaffinity chromatography using antibodies specific for EXMAD. An immunoaffinity column is constructed by covalently coupling anti-EXMAD antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing EXMAD are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of EXMAD (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/EXMAD binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and EXMAD is collected.

XV. Identification of Molecules Which Interact with EXMAD

EXMAD, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled EXMAD, washed, and any wells with labeled EXMAD complex are assayed. Data obtained using different concentrations of EXMAD are used to calculate values for the number, affinity, and association of EXMAD with the candidate molecules.

Alternatively, molecules interacting with EXMAD are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	26	398269	PITUNOT02	265928H1 (HNT2AGT01), 398269H1 and 398269R6 (PITUNOT02), 516201R6 (MMLR1DT01), 822473R6 (KERANOT02), 1265919F1 (BRAINOT09), 1356244F6 (LUNGNOT09), 1379344T6 (LUNGNOT10), 3586102H1 (293TF4T01), SBLA02091F1, SBLA01281F1
2	27	1258888	MENITUT03	1258888H1 (MENITUT03), 1373184H1 (BSTMNON02), 2420735R6 (SCORNON02), 2697827F3 (UTRSNOT12), 2990569T6 (KIDNFET02), SBCA02402F1, SBCA05599F1, SBCA01330F1, SBGA07058F3
3	28	1375891	LUNGNOT10	1375891H1 (LUNGNOT10), 2251462R6 (OVARTUT01), 4542640H1 (THYRTMT01), SAXA00188F1, SAXA00819F1, SAXA00256F1, SAXA00101F1, SZAC00197F1
4	29	1524355	UCMCL5T01	008503T6 (HMC1NOT01), 425033R6 (BLADNOT01), 1299403T6 (BRSTNOT07), 1524355H1 (UCMCL5T01), 2480893F6 (SMCANOT01), 3072568F6 (UTRSNOR01), 3077770H1 (BONEUNT01), 3521659H1 (LUNGNON03), 3810130H1 (CONTTUT01), 4187444H1 (BRSTNOT31)
5	30	1598937	BLADNOT03	307298R6 (HEARNOT01), 637901F1 (BRSTNOT03), 872833R1 (LUNGAST01), 1360462F1 (LUNGNOT12), 1598937H1 (BLADNOT03), 1688334H1 (PROSTUT10), 2048691F6 (LIVRFET02), 3604769H1 (LUNGNOT30)
6	31	1725801	PROSNOT14	359107F1 and 359107R1 (SYNORAB01), 1725801H1 and 1725801X18C1 (PROSNOT14), 2853280H1 (CONNNOT02), SBWA02129V1
7	32	1730482	BRSTTUT08	1261313R1 (SYNORAT05), 1321141F1 (BLADNOT04), 1484641F1 (CORPNOT02), 1730482H1 (BRSTTUT08), 1848053F6 (OVARNOT07), 2208990F6 (SINTFET03), 2691973F6 (LUNGNOT23), 2811183H1 (OVARNOT10), 3097712H1 (CERVNOT03), 3110665H1 (BRSTNOT17), 3738668H1 (MENTNOT01)
8	33	1810058	PROSTUT12	571697H1 (OVARNON01), 1704596F6 (DUODNOT02), 1810058H1, 1810548F6, and 1810548T6 (PROSTUT12)
9	34	2040679	HIPONON02	2040679H1 and 2040679R6 (HIPONON02), 2380160F6 (ISLTNOT01), 2621171T6 (KERANOT02), 2869976F6 (THYRNOT10)
10	35	2960051	ADRENOT09	2960051F6 and 2960051H1 (ADRENOT09), SBVA05142V1, SBVA03774V1, SBVA03935V1

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
11	36	3117318	LUNGTUT13	393775H1 (TMLR2DT01), 486988H1 (HNT2AGT01), 3117318F6 and 3117318H1 (LUNGTUT13), 3293662F6 (TLYJINT01), SBMA01131F1
12	37	3486992	EPIGNOT01	2615184H1 (GBLANOT01), 3486992H1 (EPIGNOT01), SBKA01303F1.comp, SBKA03723F1.comp, SBKA02206F1, SBKA01625F1.comp, SBKA02769F1, SBKA03712F1, SBKA02365F1, SBKA01975F1
13	38	4568384	HELATXT01	080350F1 (SYNORAB01), 320872H1 (EOSIHET02), 1418995F1 (KIDNNOT09), 1473647T1 (LUNGTUT03), 1664971F6 (BRSTNOT09), 1738547F6 (COLNNOT22), 2367046F6 (ADRENOT07), 4568384F6 and 4568384H1 (HELATXT01)
14	39	4586187	OVARNOT13	306792F1 and 306792X11R1 (HEARNOT01), 632244F1 (KIDNNOT05), 876626R1 (LUNGAST01), 2451238F6 (ENDANOT01), 2881494F6 (UTRSTUT05), 4586187H1 (OVARNOT13), 5852878H1 (FIBAUNT02), SZZZ01051R1
15	40	401801	TMLR3DT01	401801T6 and 401801H1 (TMLR3DT01), 938106H1 (CERVNOT01), 2603123T6 (UTRSNOT10), 2607556H1 (LUNGTUT07)
16	41	1721842	BLADNOT06	1721842H1, 1721842F6 and 1721842T6 (BLADNOT06), 2010387R6 (TESTNOT03), 4884119H1 (LUNLTMT01)
17	42	1833221	BRAINON01	001593H1 (U937NOT01), 389513R1 (THYMNOT02), 428370R6 (BLADNOT01), 493657H1 (HNT2NOT01), 1263824R1 (SYNORAT05), 1833221H1 (BRAINON01), 1907733F6 (CONNTUT01), 1997529R6 (BRSTTUT03), 2174658F6 (ENDCNOT03), 3114306H1 (BRSTNOT17), 3233178H1 (COLNUCT03), 4788994F6 (EPIBUNT01), 5541215H1
18	43	2041168	HIPONON02	849897R1 (NGANNOT01), 908128R2 (COLNNOT09), 999830R6 (KIDNTUT01), 1639572T6 (UTRSNOT06), 1686825F6 (PROSNOT15), 2041168H1 (HIPONON02), 2582551H1 (KIDNTUT13), 2867048H1 (KIDNNOT20), 3226063F6 (TLYJINT01), 3226063H1 (TLYJINT01), 3466031H1 (293TF2T01), 4662252H2 (BRSTTUT20), SBIA03151D1
19	44	2365794	ADRENOT07	874804H1 (LUNGAST01), 1318960T1 (BLADNOT04)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
20	45	2618452	GBLANOT01	1730514F6 (BRSTTUT08), 2225286F6 (SEMVNOT01), 2225720F6 (SEMVNOT01), 2618452F6 and 2618452H1 (GBLANOT01), 2618457F6 (GBLANOT01), 3248134H1 (SEMVNOT03), 3250560H1 (SEMVNOT03), 3538176F6 (SEMVNOT04), 4068913H1 (SEMVNOT05)
21	46	2622288	KERANOT02	223636F1 (PANCNOT01), 490914R6 (HNT2AGT01), 530368R6 (BRAINOT03), 850583R1 (NGANNOT01), 898618R1 (BRSTTUT03), 932484R6 (CERVNOT01), 1302418F1 (PLACNOT02), 1368735R1 (SCORNON02), 1486177F6 (CORPNOT02), 1726367F6 (PROSNOT14), 2516869H1 (LIVRTUT04), 2622288R6 and 2622288H2 (KERANOT02), 3043955H1 (HEAANOT01), 3398316H1 (UTRSNOT16), 3938796H1 (SKINBIT01), 4043471H1 (LUNGNOT35)
22	47	2806595	BLADTUT08	643445R6 (BRSTTUT02), 2806595F6 and 2806595H1 (BLADTUT08), SBRA04014D1, SBRA03510D1
23	48	2850987	BRSTTUT13	1300925F1 (BRSTNOT07), 1339833F1 (COLNTUT03), 1347463F6 (PROSNOT11), 1347463T6 (PROSNOT11), 1899642F6 (BLADTUT06), 2715093F6 (THYRNOT09), 2726463F6 (OVARTUT05), 2850987H1 (BRSTTUT13), 2893008H1 (LUNGFET04), 3336701F6 (SPLNNOT10), 3341661H1 (SPLNNOT09), SXAF00652V1, SXAF03272V1
24	49	3557211	LUNGNOT31	958552H1 (KIDNNOT05), 2953281F6 and 2953281T6 (KIDNFET01), 3557211F6 and 3557211H1 (LUNGNOT31), 4306204H1 (GBLADIT01), 4420950F6 (LIVRDIT02) g2188176, g1424165
25	50	4675668	NOSEDIT02	1519431T6 (BLADTUT04), 2447058F6 (THP1NOT03), 2758306R6 (THP1AZS08), 2758306T6 (THP1AZS08), 3589494H1 (293TF5T01), 3813434H1 (TONSNOT03), 4675668H1 (NOSEDIT02), 5175727H1 (EPIBTXT01), 5313381H1 (KIDETXS02)

Table 2

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
1	309	T153 S29 S140 T153 S162 T168 S233 S258 T285 S290 T87 T159 T265	N108 N305	Signal peptide: M1-A31	similar to B. Subtilis surfactin (SFP) protein g3880360	BLAST SPSCAN
2	554	S57 S146 S265 T275 S389 T495 T496 S497 S551 S25 S34 T87 S115 S180 S212 S242 S289 T308 S361 T388 T504	N398	EGF-like domain: C98-C132 C138-C172 C178-C217 C223-C258 Cell adhesion: R363-D365 Signal peptide: M1-G21	fibulin-2 [Mus musculus] g437047	BLAST PRINTS BLOCKS PFAM MOTIFS SPSCAN HMM
3	482	S87 T37 T108 T131 S133 S148 T165 T246 S254 T256 S269 S283 S333 S404 T414 T431 S28 T29 S65 T335 T431 S446 S460 T464	N252 N445 N451	Signal peptide: M1-G22	gastric mucin [Sus scrofa] g915208	BLAST MOTIFS SPSCAN HMM
4	735	S506 S153 S243 T259 S304 T317 T378 S414 T502 S575 S670 S688 S698 S44 T116 S258 S324 S350 S356 S396 T437 T515 S610 S620 Y53	N70 N97 N144 N188 N412	Kelch motif: T284-K330 C469-G513	muskelin [Mus musculus] g3493462	BLAST PFAM

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
5	424	T209 S256 S276 T86 S311 S319 T347 S15 S354 S394 S107 Y53 S153 T217 S258 S408 T306 S358 S383		SH3 domain: V366-V422	Focal adhesion protein (FAP52) [Gallus gallus] g2217964	BLAST PFAM PRINTS BLOCKS
6	420	S293 T63 T73 S99 S101 S222 T359 T48 T63 S138 T159 S406 S409 Y53	N79 N205	Signal peptide: M1-L29 EGF-like domain: T174-C192 Cysteine-rich pattern: C181-C192	HT protein [Cricetulus griseus] g1216486	BLAST PRINTS SPSCAN MOTIFS HMM
7	795	S41 T94 S145 S243 T297 S442 S451 T687 S103 T111 T129 S184 T428 S647	N383 N387	Cell adhesion: R606-D608 von Willebrand factor type A domain: D31-L204 transmembrane domain: I50-T77	collagen type XIV [Homo sapiens] g2065167	BLAST MOTIFS PFAM PRINTS HMM
8	306	T69 T133 S255 T279 T22		Signal peptide: M1-S19 Clq domain: G149-P175 A203-I226 H227-L302	saccular collagen [Lepomis macrochirus] g687606	BLAST PFAM PRINTS BLOCKS SPSCAN HMM

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
9	338	S5 S53 S66 T119 T246 S23 T65 S102 S151 S251 T277	N217 N332	Signal peptide: M1-S22 Leucine-rich repeats domain: S102-T147 S151-I196 N197-A243	LRR47 [Drosophila melanogaster] g415947	BLAST PFAM PRINTS SPSCAN HMM
10	164	S42 S75 T160 S44 S49		Signal peptide: M1-G20 von Willebrand factor C-type domain: C103-C157	extracellular matrix protein [Homo sapiens] g3786312	BLAST PFAM BLOCKS SPSCAN HMM
11	327	S292 S30 S35 S63 T92 T14 T102 T179 S198 T285	N54 N61 N75 N85 N100 N189 N196 N213 N218 N322	Signal peptide: M1-P29 Ig domain: P81-F144 G173-A239 Transmembrane domain: V254-A276	embigin protein [Rattus norvegicus] g3355709	BLAST PFAM SPSCAN HMM
12	716	S21 T49 T54 T87 T98 S245 T315 T471 T519 T590 S624 S692 T705 S176 S384 S473 S494 T513 S542 T560 T571 T605 T613 S664 T709 Y581	N69 N96 N106 N117 N385 N517 N582 N611	Signal peptide: M1-S25 Leucine-rich repeats domain: N96-S143 N192-D239 S240-L287 R288-P337 A338-N385 Transmembrane domain: M639-F656	leucine-rich-repeat protein [Mus musculus] g1228052	BLAST PFAM PRINTS SPSCAN HMM

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
13	665	T147 S45 S86 S110 S121 T147 S160 T200 S205 S225 S247 S299 S301 S309 S335 S336 S341 S343 T386 S388 T400 T448 S506 S534 S545 S580 S581 S582 S597 S602 S615 S23 S82 S100 S162 S183 T199 S217 S221 S329 S347 T429 T501 T558 T563 T608 Y445 Y559	N119 N242 N424 N427 N634		50kDa lectin [Bombyx mori] g500858	BLAST
14	547	T60 S31 T87 T175 S213 T357 T452 T474 S476 T488 S203 T420 Y424	N15 N76 N85 N104 N128 N154 N191 N221 N242 N418	Lectin C-type domain: L473-C535 T488-L547 Cell adhesion: R256-D258	CSR1 (cellular stress response protein) [Homo sapiens] g6230372	BLAST PFAM BLOCKS MOTIFS PPROFILESCAN
15	109	S85 S38	N22		Attractin; DPPT-L [Homo sapiens] g3676347	BLAST-GenBank MOTIFS
16	192	S10 S87 T92 T157 T165 T170 S19 S46	N8 N103	Leucine Rich Repeat Domain: L81-I94 L126-M139		BLIMPS-PRINTS MOTIFS

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
17	575	T150 S171 S299 S85 S98 S117 S118 S126 S142 S170 S203 S237 S239 S333 S415 S467 T473 S524 T557 S558 S562 S32 S92 S104 S128 S134 T149 T150 S167 S188 S260 S270 S280 S289 S389 S536	N68 N96 N234 N366 N569		axotrophin [Mus musculus] g5052031 dentin phosphoryn [Homo sapiens] g4322670	BLAST-GenBank MOTIFS
18	342	S73 S24 S82 S207 S315 S96 T176	N31 N152 N180 N193	Armadillo/beta-catenin-like repeats: A104-A113		BLIMPS-PFAM MOTIFS
19	110	S80		Signal Peptide: M1-G45 Transmembrane Domain: G48-G71 G91-Y110 Legume lectins signature: V4-F54		SPSCAN HMMER PROFILESCAN MOTIFS
20	571	S482 T502 T11 T40 S88 T180 S231 T339 T383 T402 T409 T436 T447 S482 T491	N66 N229 N434 N498	Mucin domain: P101 - S430 Cystine knot domain: C481-C569	mucin [Homo sapiens] g292046	BLAST-GenBank BLAST-DOMO HMMER-PFAM MOTIFS

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
21	262	S69 S146 S172 S41 T54 T59 T101 T102 T107 Y170		Signal Peptide: M1-G25	single-pass transmembrane protein [Rattus norvegicus] g6978944 antigen [Homo sapiens] g188543	SPSCAN HMMER MOTIFS BLAST-GenBank
22	172	S29 T53 S111 S80 Y144		Signal Peptide: M1-G17 Protein proteoglycan core glycoprotein precursor cartilage repeat lectin Ig fold : G63-I149 Immunoglobulin: E52-S156	link protein [Mus musculus] g4218976	BLAST-GenBank BLAST-PRODOM BLAST-DOMO SPSCAN HMMER MOTIFS
23	571	S16 T36 T294 S396 S403 T445 S23 T176 S487	N100 N174 N434 N567	Mitochondrial energy transfer proteins signature: P404-F412 Transmembrane domains: T94-K116 F520-F539 L58-I78 I341-W362 I375-M393 I453-F472 Laminin b: Y538-K558	cell adhesion regulator [Rattus norvegicus] g4098299	BLAST-GenBank, HMMER-PFAM HMMER MOTIFS

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequence	Homologous Sequence	Analytical Methods
24	455	S18 S23 S143 S270 S81 T186 S196 T208 S230 T240 T256 S418 S452 Y223	N138 N217 N288	Signal peptidases I signature: G43-F50 Lectin c-type: C329-S452 Cell attachment sequence: R183-D185	lectin BRA-3 [Megabalanus rosa] g407227	BLAST-GenBank BLAST-DOMO HMMER-PFAM MOTIFS
25	437	S98 T146 T160 S211 T220 T301 S55 T86 T156 S197 T369 Y265 Y334 Y350		ENP1 protein nuclear protein: E157-D431	bystin [Mus musculus] g2738509	BLAST-GenBank BLAST-PRODOM MOTIFS

Table 3

Nucleotide SEQ ID NO:	Unique Fragment	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
26	242-286	Nervous (0.264) Reproductive (0.198)	Cancer (0.462) Cell proliferation (0.242) Inflammation (0.176)	PSPORT
27	272-316	Nervous (0.438) Reproductive (0.188) Developmental (0.188)	Cancer (0,438) Cell proliferation (0.250) Inflammation (0.188)	pINCY
28	218-262	Gastrointestinal (0.244) Nervous (0.195) Reproductive (0.171)	Cancer (0.488) Inflammation (0.195) Cell proliferation (0.146)	pINCY
29	488-532 1082-1126	Reproductive (0.265) Nervous (0.206) Hematopoietic/immune (0.147)	Cancer (0.500) Cell proliferation (0.324) Inflammation (0.235)	PBLUESCRIPT
30	542-586	Reproductive (0.321) Cardiovascular (0.143) Musculoskeletal (0.143)	Cancer (0.500) Inflammation (0.107) Cell proliferation (0.107)	pINCY
31	217-261	Nervous (0.265) Reproductive (0.253) Cardiovascular (0.108)	Cancer (0.482) Inflammation (0.145) Cell proliferation (0.145)	pINCY
32	36-80	Reproductive (0.333) Gastrointestinal (0.154) Developmental (0.115)	Cancer (0.462) Inflammation (0.167) Cell proliferation (0.154)	pINCY
33	218-262	Reproductive (0.571) Gastrointestinal (0.286) Cardiovascular (0.143)	Trauma (0.286) Cancer (0.143) Inflammation (0.143)	pINCY
34	111-155	Gastrointestinal (0.364) Nervous (0.182) Cardiovascular (0.091)	Cancer (0.364) Inflammation (0.273) Cell proliferation (0.182)	PSPORT
35	271-315	Musculoskeletal (0.286) Reproductive (0.286) Cardivascular (0.143)	Cancer (0.286) Inflammation (0.143) Neurological (0.143)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Unique Fragment	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
36	542-586 866-910	Hematopoietic/Immune (0.526) Reproductive (0.158) Nervous (0.105)	Cancer (0.368) Inflammation (0.474) Cell proliferation (0.158)	pINCY
37	811-855	Nervous (0.267) Reproductive (0.267) Musculoskeletal (0.133)	Cancer (0.600) Inflammation (0.200) Cell proliferation (0.133)	pINCY
38	380-424 974-1018	Reproductive (0.200) Gastrointestinal (0.164) Nervous (0.145)	Cancer (0.436) Cell proliferation (0.309) Inflammation (0.200)	pINCY
39	434-479 975-1019	Reproductive (0.296) Cardiovascular (0.259) Hematopoeitic/Immune (0.111)	Cancer (0.315) Inflammation (0.204) Trauma (0.204)	pINCY
40	555-614	Cardiovascular (0.333) Hematopoietic/Immune (0.333) Reproductive (0.333)	Inflammation (0.667) Cancer (0.333)	PBLUESCRIPT
41	743-802	Nervous (0.353) Reproductive (0.176) Urologic (0.176)	Cancer (0.471) Inflammation (0.411) Cell Proliferation (0.118)	pINCY
42	429-488 1029-1088	Reproductive (0.213) Nervous (0.191) Cardiovascular (0.169)	Cancer (0.472) Inflammation (0.394) Cell Proliferation (0.180)	PSPORT1
43	967-1026	Nervous (0.228) Reproductive (0.213) Gastrointestinal (0.110)	Cancer (0.504) Inflammation (0.291) Cell Proliferation (0.197)	PSPORT1
44	164-223	Reproductive (0.241) Cardiovascular (0.167) Gastrointestinal (0.148)	Cancer (0.481) Inflammation (0.315) Cell Proliferation (0.167)	pINCY
45	110-169	Gastrointestinal (0.562) Reproductive (0.312) Nervous (0.062) Urologic (0.062)	Cancer (0.500) Inflammation (0.312) Cell Proliferation (0.062)	pINCY

Table 3 (cont.)

Nucleotide SEQ ID NO:	Unique Fragment	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
46	273-332 759-818	Nervous (0.347) Reproductive (0.223) Cardiovascular (0.132)	Cancer (0.430) Inflammation (0.364) Cell Proliferation (0.124)	PSPORT1
47	218-277	Gastrointestinal (0.200) Nervous (0.200) Reproductive (0.200)	Cancer (0.533) Inflammation (0.334) Cell Proliferation (0.133)	pINCY
48	341-400	Reproductive (0.294) Gastrointestinal (0.168) Cardiovascular (0.126)	Cancer (0.476) Inflammation (0.329) Cell Proliferation (0.168)	pINCY
49	266-325 542-601	Developmental (0.277) Gastrointestinal (0.222) Nervous (0.167) Urologic (0.167)	Cell Proliferation (0.444) Inflammation (0.444) Cancer (0.167)	pINCY
50	165-224	Hematopoietic/Immune (0.216) Reproductive (0.216) Gastrointestinal (0.135)	Cancer (0.568) Cell Proliferation (0.324) Inflammation (0.297)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Description
26	PITUNOT02	The library was constructed using RNA obtained from Clontech. The RNA was isolated from the pituitary glands removed from a pool of 87 male and female donors, 15 to 75 years old.
27	MENITUT03	The library was constructed using RNA isolated from brain meningioma tissue removed from a 35-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a benign neoplasm in the right cerebellopontine angle of the brain. Patient history included hypothyroidism. Family history included myocardial infarction and breast cancer.
28	LUNGNOT10	The library was constructed using RNA isolated from the lung tissue of a Caucasian male fetus, who died at 23 weeks' gestation.
29	UCMCL5T01	The UCMCL5T01 library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates.
30	BLADNOT03	The library was constructed using RNA isolated from the bladder tissue removed from an 80-year-old Caucasian female during a radical cystectomy and lymph node excision. Pathology for the associated tumor tissue indicated grade 3 invasive transitional cell carcinoma. Patient history included malignant neoplasm of the uterus, atherosclerosis, and atrial fibrillation. Family history included acute renal failure, osteoarthritis, and atherosclerosis.
31	PROSNOT14	The library was constructed using RNA isolated from diseased prostate tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst and hematuria. Family history included benign hypertension, cerebrovascular disease, and arteriosclerotic coronary artery disease.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
32	BRSTTUT08	The library was constructed using RNA isolated from breast tumor tissue removed from a 45-year-old Caucasian female during unilateral extended simple mastectomy. Pathology indicated invasive nuclear grade 2-3 adenocarcinoma, ductal type, with 3 of 23 lymph nodes positive for metastatic disease. Greater than 50% of the tumor volume was in situ, both comedo and non-comedo types. Immunostains were positive for estrogen/progesterone receptors, and uninvolved tissue showed proliferative changes. The patient concurrently underwent a total abdominal hysterectomy. Patient history included valvuloplasty of mitral valve without replacement, rheumatic mitral insufficiency, and rheumatic heart disease. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
33	PROSTUT12	The library was constructed using RNA isolated from prostate tumor tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology indicated an adenocarcinoma (Gleason grade 2+2). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA).
34	HIPONON02	This normalized hippocampus library was constructed from 1.13M independent clones from a normal hippocampus library. RNA was isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis. The normalization and hybridization conditions were adapted from Soares et al. (Proc.Natl.Acad.Sci. USA (1994) 91:9228).
35	ADRENOT09	The library was constructed using RNA isolated from left adrenal gland tissue removed from a 43-year-old Caucasian male during nephroureterectomy, regional lymph node excision, and unilateral left adrenalectomy. Pathology for the associated tumor tissue indicated a grade 2 renal cell carcinoma mass in the posterior lower pole of the left kidney with invasion into the renal pelvis.
36	LUNGTUT13	The library was constructed using RNA isolated from tumorous lung tissue removed from the right upper lobe of a 47-year-old Caucasian male during a segmental lung resection. Pathology indicated invasive grade 3 (of 4) adenocarcinoma. Family history included atherosclerotic coronary artery disease, and type II diabetes.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
37	EPIGNOT01	The library was constructed using RNA isolated from epiglottic tissue removed from a 71-year-old male during laryngectomy with right parathyroid biopsy. Pathology for the associated tumor tissue indicated recurrent grade 1 papillary thyroid carcinoma.
38	HELATXT01	The library was constructed using RNA isolated from HeLa cells treated with TNF-a and IL-1b, 10ng/nl each for 20 hours. The HeLa cell line is derived from cervical adenocarcinoma removed from a 31-year-old Black female.
39	OVARNOT13	The library was constructed using RNA isolated from left ovary tissue removed from a 47-year-old Caucasian female during a vaginal hysterectomy with bilateral salpingo-oophorectomy, and dilation and curettage. Pathology for the associated tumor tissue indicated a single intramural leiomyoma. The endometrium was in the secretory phase. The patient presented with metrorrhagia. Patient history included hyperlipidemia and benign hypertension. Family history included colon cancer, benign hypertension, atherosclerotic coronary artery disease, and breast cancer.
40	TMLR3DT01	Library was constructed using RNA isolated from non-adherent and adherent peripheral blood mononuclear cells collected from two unrelated Caucasian male donors (25 and 29 years old). Cells from each donor were purified on Ficoll Hypaque, then co-cultured for 96 hours in medium containing normal human serum at a cell density of 2×10^6 cells/ml. The non-adherent and adherent cell populations were pooled, washed once in PBS, lysed in a buffer containing GuSCN, and spun through CsCl to obtain RNA.
41	BLADNOT06	Library was constructed using RNA isolated from the posterior wall bladder tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy and urinary diversion. Pathology for the associated tumor tissue indicated grade 3 transitional cell carcinoma on the anterior wall of the bladder and urothelium. Patient history included lung neoplasm, and tobacco abuse in remission. Family history included a malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
42	BRAINON01	Library was constructed and normalized from 4.88 million independent clones from a brain library. RNA was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
43	HIPONON02	This normalized hippocampus library was constructed from 1.13 million independent clones from a hippocampal library. RNA was isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis. The normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9928).
44	ADRENOT07	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
45	GBLANOT01	Library was constructed using RNA isolated from diseased gallbladder tissue removed from a 53-year-old Caucasian female during a cholecystectomy. Pathology indicated mild chronic cholecystitis and cholelithiasis with approximately 150 mixed gallstones. Family history included benign hypertension.
46	KERANOT02	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
47	BLADTUT08	Library was constructed using RNA isolated from bladder tumor tissue removed from a 72-year-old Caucasian male during a radical cystectomy and prostatectomy. Pathology indicated an invasive grade 3 (of 3) transitional cell carcinoma in the right bladder base. Patient history included pure hypercholesterolemia and tobacco abuse. Family history included myocardial infarction, cerebrovascular disease, brain cancer, and myocardial infarction.
48	BRSTTUT13	Library was constructed using RNA isolated from breast tumor tissue removed from the right breast of a 46-year-old Caucasian female during a unilateral extended simple mastectomy with breast reconstruction. Pathology indicated an invasive grade 3 adenocarcinoma, ductal type with apocrine features and greater than 50% intraductal component. Patient history included breast cancer.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Description
49	LUNGNOT31	Library was constructed using RNA isolated from right middle lobe lung tissue removed from a 63-year-old Caucasian male. Pathology for the associated tumor indicated grade 3 adenocarcinoma. Patient history included an abdominal aortic aneurysm, cardiac dysrhythmia, atherosclerotic coronary artery disease, hiatal hernia, chronic sinusitis, and lupus. Family history included acute myocardial infarction and atherosclerotic coronary artery disease.
50	NOSEDIT02	The library was constructed using RNA isolated from nasal polyp tissue.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-25,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-25.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide of claim 3 selected from the group consisting of SEQ ID NO:26-50.

5. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

6. A cell transformed with a recombinant polynucleotide of claim 5.

7. A transgenic organism comprising a recombinant polynucleotide of claim 5.

8. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

9. An isolated antibody which specifically binds to a polypeptide of claim 1.

10. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50,

b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

11. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 10.

12. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 10, the method comprising:

a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

13. A method of claim 12, wherein the probe comprises at least 30 contiguous nucleotides.

14. A method of claim 12, wherein the probe comprises at least 60 contiguous nucleotides.

15. A pharmaceutical composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

16. A method for treating a disease or condition associated with decreased expression of functional EXMAD, comprising administering to a patient in need of such treatment the pharmaceutical composition of claim 15.

17. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

18. A pharmaceutical composition comprising an agonist compound identified by a method of claim 17 and a pharmaceutically acceptable excipient.

19. A method for treating a disease or condition associated with decreased expression of functional EXMAD, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 18.

20. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

21. A pharmaceutical composition comprising an antagonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with overexpression of functional EXMAD, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 21.

23. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 4, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
BANDMAN, Olga
HILLMAN, Jennifer L.
TANG, Y. Tom
LAL, Preeti
YUE, Henry
BAUGHN, Mariah R.
LU, Dyung Aina M.
AZIMZAI, Yalda

<120> EXTRACELLULAR MATRIX AND ADHESION-ASSOCIATED PROTEINS

<130> PF-0693 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/133,643; 60/150,409
<151> 1999-05-11; 1999-08-23

<160> 50
<170> PERL Program

<210> 1
<211> 309
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 398269CD1

<400> 1
```
Met Val Phe Pro Ala Lys Arg Phe Cys Leu Val Pro Ser Met Glu
  1               5                  10                  15
Gly Val Arg Trp Ala Phe Ser Cys Gly Thr Trp Leu Pro Ser Arg
                 20                  25                  30
Ala Glu Trp Leu Leu Ala Val Arg Ser Ile Gln Pro Glu Glu Lys
                 35                  40                  45
Glu Arg Ile Gly Gln Phe Val Phe Ala Arg Asp Ala Lys Ala Ala
                 50                  55                  60
Met Ala Gly Arg Leu Met Ile Arg Lys Leu Val Ala Glu Lys Leu
                 65                  70                  75
Asn Ile Pro Trp Asn His Ile Arg Leu Gln Arg Thr Ala Lys Gly
                 80                  85                  90
Lys Pro Val Leu Ala Lys Asp Ser Ser Asn Pro Tyr Pro Asn Phe
                 95                 100                 105
Asn Phe Asn Ile Ser His Gln Gly Asp Tyr Ala Val Leu Ala Ala
                110                 115                 120
Glu Pro Glu Leu Gln Val Gly Ile Asp Ile Met Lys Thr Ser Phe
                125                 130                 135
Pro Gly Arg Gly Ser Ile Pro Glu Phe Phe His Ile Met Lys Arg
                140                 145                 150
Lys Phe Thr Asn Lys Glu Trp Glu Thr Ile Arg Ser Phe Lys Asp
                155                 160                 165
Glu Trp Thr Gln Leu Asp Met Phe Tyr Arg Asn Trp Ala Leu Lys
                170                 175                 180
Glu Ser Phe Ile Lys Ala Ile Gly Val Gly Leu Gly Phe Glu Leu
                185                 190                 195
Gln Arg Leu Glu Phe Asp Leu Ser Pro Leu Asn Leu Asp Ile Gly
                200                 205                 210
Gln Val Tyr Lys Glu Thr Arg Leu Phe Leu Asp Gly Glu Glu Glu
                215                 220                 225
Lys Glu Trp Ala Phe Glu Glu Ser Lys Ile Asp Glu His His Phe
```

230 235 240
Val Ala Val Ala Leu Arg Lys Pro Asp Gly Ser Arg His Gln Asp
245 250 255
Val Pro Ser Gln Asp Asp Ser Lys Pro Thr Gln Arg Gln Phe Thr
260 265 270
Ile Leu Asn Phe Asn Asp Leu Met Ser Ser Ala Val Pro Met Thr
275 280 285
Pro Glu Asp Pro Ser Phe Trp Asp Cys Phe Cys Phe Thr Glu Glu
290 295 300
Ile Pro Ile Arg Asn Gly Thr Lys Ser
305

<210> 2
<211> 554
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1258888CD1

<400> 2
Met Pro Leu Pro Trp Ser Leu Ala Leu Pro Leu Leu Leu Ser Trp
1 5 10 15
Val Ala Gly Gly Phe Gly Asn Ala Ala Ser Ala Arg His His Gly
20 25 30
Leu Leu Ala Ser Ala Arg Gln Pro Gly Val Cys His Tyr Gly Thr
35 40 45
Lys Leu Ala Cys Cys Tyr Gly Trp Arg Arg Asn Ser Lys Gly Val
50 55 60
Cys Glu Ala Thr Cys Glu Pro Gly Cys Lys Phe Gly Glu Cys Val
65 70 75
Gly Pro Asn Lys Cys Arg Cys Phe Pro Gly Tyr Thr Gly Lys Thr
80 85 90
Cys Ser Gln Asp Val Asn Glu Cys Gly Met Lys Pro Arg Pro Cys
95 100 105
Gln His Arg Cys Val Asn Thr His Gly Ser Tyr Lys Cys Phe Cys
110 115 120
Leu Ser Gly His Met Leu Met Pro Asp Ala Thr Cys Val Asn Ser
125 130 135
Arg Thr Cys Ala Met Ile Asn Cys Gln Tyr Ser Cys Glu Asp Thr
140 145 150
Glu Glu Gly Pro Gln Cys Leu Cys Pro Ser Ser Gly Leu Arg Leu
155 160 165
Ala Pro Asn Gly Arg Asp Cys Leu Asp Ile Asp Glu Cys Ala Ser
170 175 180
Gly Lys Val Ile Cys Pro Tyr Asn Arg Arg Cys Val Asn Thr Phe
185 190 195
Gly Ser Tyr Tyr Cys Lys Cys His Ile Gly Phe Glu Leu Gln Tyr
200 205 210
Ile Ser Gly Arg Tyr Asp Cys Ile Asp Ile Asn Glu Cys Thr Met
215 220 225
Asp Ser His Thr Cys Ser His His Ala Asn Cys Phe Asn Thr Gln
230 235 240
Gly Ser Phe Lys Cys Lys Cys Lys Gln Gly Tyr Lys Gly Asn Gly
245 250 255
Leu Arg Cys Ser Ala Ile Pro Glu Asn Ser Val Lys Glu Val Leu
260 265 270
Arg Ala Pro Gly Thr Ile Lys Asp Arg Ile Lys Lys Leu Leu Ala
275 280 285
His Lys Asn Ser Met Lys Lys Lys Ala Lys Ile Lys Asn Val Thr
290 295 300
Pro Glu Pro Thr Arg Thr Pro Thr Pro Lys Val Asn Leu Gln Pro
305 310 315
Phe Asn Tyr Glu Glu Ile Val Ser Arg Gly Gly Asn Ser His Gly
320 325 330
Gly Lys Lys Gly Asn Glu Glu Lys Met Lys Glu Gly Leu Glu Asp
335 340 345

Glu Lys Arg Glu Glu Lys Ala Leu Lys Asn Asp Ile Glu Glu Arg
350 355 360
Ser Leu Arg Gly Asp Val Phe Phe Pro Lys Val Asn Glu Ala Gly
365 370 375
Glu Phe Gly Leu Ile Leu Val Gln Arg Lys Ala Leu Thr Ser Lys
380 385 390
Leu Glu His Lys Ala Asp Leu Asn Ile Ser Val Asp Cys Ser Phe
395 400 405
Asn His Gly Ile Cys Asp Trp Lys Gln Asp Arg Glu Asp Asp Phe
410 415 420
Asp Trp Asn Pro Ala Asp Arg Asp Asn Ala Ile Gly Phe Tyr Met
425 430 435
Ala Val Pro Ala Leu Ala Gly His Lys Lys Asp Ile Gly Arg Leu
440 445 450
Lys Leu Leu Leu Pro Asp Leu Gln Pro Gln Ser Asn Phe Cys Leu
455 460 465
Leu Phe Asp Tyr Arg Leu Ala Gly Asp Lys Val Gly Lys Leu Arg
470 475 480
Val Phe Val Lys Asn Ser Asn Asn Ala Leu Ala Trp Glu Lys Thr
485 490 495
Thr Ser Glu Asp Glu Lys Trp Lys Thr Gly Lys Ile Gln Leu Tyr
500 505 510
Gln Gly Thr Asp Ala Thr Lys Ser Ile Ile Phe Glu Ala Glu Arg
515 520 525
Gly Lys Gly Lys Thr Gly Glu Ile Ala Val Asp Gly Val Leu Leu
530 535 540
Val Ser Gly Leu Cys Pro Asp Ser Leu Leu Ser Val Asp Asp
545 550

<210> 3
<211> 482
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1375891CD1

<400> 3
Met Gly Cys Leu Trp Gly Leu Ala Leu Pro Leu Phe Phe Phe Cys
1 5 10 15
Trp Glu Val Gly Val Ser Gly Ser Ser Ala Gly Pro Ser Thr Arg
20 25 30
Arg Ala Asp Thr Ala Met Thr Thr Asp Asp Thr Glu Val Pro Ala
35 40 45
Met Thr Leu Ala Pro Gly His Ala Ala Leu Glu Thr Gln Thr Leu
50 55 60
Ser Ala Glu Thr Ser Ser Arg Ala Ser Thr Pro Ala Gly Pro Ile
65 70 75
Pro Glu Ala Glu Thr Arg Gly Ala Lys Arg Ile Ser Pro Ala Arg
80 85 90
Glu Thr Arg Ser Phe Thr Lys Thr Ser Pro Asn Phe Met Val Leu
95 100 105
Ile Ala Thr Ser Val Glu Thr Ser Ala Ala Ser Gly Ser Pro Glu
110 115 120
Gly Ala Gly Met Thr Thr Val Gln Thr Ile Thr Gly Ser Asp Pro
125 130 135
Glu Glu Ala Ile Phe Asp Thr Leu Cys Thr Asp Asp Ser Ser Glu
140 145 150
Glu Ala Lys Thr Leu Thr Met Asp Ile Leu Thr Leu Ala His Thr
155 160 165
Ser Thr Glu Ala Lys Gly Leu Ser Ser Glu Ser Ser Ala Ser Ser
170 175 180
Asp Gly Pro His Pro Val Ile Thr Pro Ser Arg Ala Ser Glu Ser
185 190 195
Ser Ala Ser Ser Asp Gly Pro His Pro Val Ile Thr Pro Ser Arg
200 205 210
Ala Ser Glu Ser Ser Ala Ser Ser Asp Gly Pro His Pro Val Ile

```
                215                     220                     225
Thr Pro Ser Trp Ser Pro Gly Ser Asp Val Thr Leu Leu Ala Glu
                230                     235                     240
Ala Leu Val Thr Val Thr Asn Ile Glu Val Ile Asn Cys Ser Ile
                245                     250                     255
Thr Glu Ile Glu Thr Thr Thr Ser Ser Ile Pro Gly Ala Ser Asp
                260                     265                     270
Ile Asp Leu Ile Pro Thr Glu Gly Val Lys Ala Ser Ser Thr Ser
                275                     280                     285
Asp Pro Pro Ala Leu Pro Asp Ser Thr Glu Ala Lys Pro His Ile
                290                     295                     300
Thr Glu Val Thr Ala Ser Ala Glu Thr Leu Ser Thr Ala Gly Thr
                305                     310                     315
Thr Glu Ser Ala Ala Pro His Ala Thr Val Gly Thr Pro Leu Pro
                320                     325                     330
Thr Asn Ser Ala Thr Glu Arg Glu Val Thr Ala Pro Gly Ala Thr
                335                     340                     345
Thr Leu Ser Gly Ala Leu Val Thr Val Ser Arg Asn Pro Leu Glu
                350                     355                     360
Glu Thr Ser Ala Leu Ser Val Glu Thr Pro Ser Tyr Val Lys Val
                365                     370                     375
Ser Gly Ala Ala Pro Val Ser Ile Glu Ala Gly Ser Ala Val Gly
                380                     385                     390
Lys Thr Thr Ser Phe Ala Gly Ser Ser Ala Ser Ser Tyr Ser Pro
                395                     400                     405
Ser Glu Ala Ala Leu Lys Asn Phe Thr Pro Ser Glu Thr Pro Thr
                410                     415                     420
Met Asp Ile Ala Thr Lys Gly Pro Phe Pro Thr Ser Arg Asp Pro
                425                     430                     435
Leu Pro Ser Val Pro Pro Thr Thr Thr Asn Ser Ser Arg Gly Thr
                440                     445                     450
Asn Ser Thr Leu Ala Lys Ile Thr Thr Ser Ala Lys Thr Thr Met
                455                     460                     465
Lys Pro Gln Gln Pro Arg Pro Arg Leu Pro Gly Arg Gly Arg Pro
                470                     475                     480
Gln Thr

<210> 4
<211> 735
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1524355CD1

<400> 4
Met Ala Ala Gly Gly Ala Val Ala Ala Ala Pro Glu Cys Arg Leu
 1               5                  10                  15
Leu Pro Tyr Ala Leu His Lys Trp Ser Ser Phe Ser Ser Thr Tyr
                20                  25                  30
Leu Pro Glu Asn Ile Leu Val Asp Lys Pro Asn Asp Gln Ser Ser
                35                  40                  45
Arg Trp Ser Ser Glu Ser Asn Tyr Pro Pro Gln Tyr Leu Ile Leu
                50                  55                  60
Lys Leu Glu Arg Pro Ala Ile Val Gln Asn Ile Thr Phe Gly Lys
                65                  70                  75
Tyr Glu Lys Thr His Val Cys Asn Leu Lys Lys Phe Lys Val Phe
                80                  85                  90
Gly Gly Met Asn Glu Glu Asn Met Thr Glu Leu Leu Ser Ser Gly
                95                 100                 105
Leu Lys Asn Asp Tyr Asn Lys Glu Thr Phe Thr Leu Lys His Lys
               110                 115                 120
Ile Asp Glu Gln Met Phe Pro Cys Arg Phe Ile Lys Ile Val Pro
               125                 130                 135
Leu Leu Ser Trp Gly Pro Ser Phe Asn Phe Ser Ile Trp Tyr Val
               140                 145                 150
```

Glu Leu Ser Gly Ile Asp Asp Pro Asp Ile Val Gln Pro Cys Leu
155 160 165
Asn Trp Tyr Ser Lys Tyr Arg Glu Gln Glu Ala Ile Arg Leu Cys
170 175 180
Leu Lys His Phe Arg Gln His Asn Tyr Thr Glu Ala Phe Glu Ser
185 190 195
Leu Gln Lys Lys Thr Lys Ile Ala Leu Glu His Pro Met Leu Thr
200 205 210
Asp Ile His Asp Lys Leu Val Leu Lys Gly Asp Phe Asp Ala Cys
215 220 225
Glu Glu Leu Ile Glu Lys Ala Val Asn Asp Gly Leu Phe Asn Gln
230 235 240
Tyr Ile Ser Gln Gln Glu Tyr Lys Pro Arg Trp Ser Gln Ile Ile
245 250 255
Pro Lys Ser Thr Lys Gly Asp Gly Glu Asp Asn Arg Pro Gly Met
260 265 270
Arg Gly Gly His Gln Met Val Ile Asp Val Gln Thr Glu Thr Val
275 280 285
Tyr Leu Phe Gly Gly Trp Asp Gly Thr Gln Asp Leu Ala Asp Phe
290 295 300
Trp Ala Tyr Ser Val Lys Glu Asn Gln Trp Thr Cys Ile Ser Arg
305 310 315
Asp Thr Glu Lys Glu Asn Gly Pro Ser Ala Arg Ser Cys His Lys
320 325 330
Met Cys Ile Asp Ile Gln Arg Arg Gln Ile Tyr Thr Leu Gly Arg
335 340 345
Tyr Leu Asp Ser Ser Val Arg Asn Ser Lys Ser Leu Lys Ser Asp
350 355 360
Phe Tyr Arg Tyr Asp Ile Asp Thr Asn Thr Trp Met Leu Leu Ser
365 370 375
Glu Asp Thr Ala Ala Asp Gly Gly Pro Lys Leu Val Phe Asp His
380 385 390
Gln Met Cys Met Asp Ser Glu Lys His Met Ile Tyr Thr Phe Gly
395 400 405
Gly Arg Ile Leu Thr Cys Asn Gly Ser Val Asp Asp Ser Arg Ala
410 415 420
Ser Glu Pro Gln Phe Ser Gly Leu Phe Ala Phe Asn Cys Gln Cys
425 430 435
Gln Thr Trp Lys Leu Leu Arg Glu Asp Ser Cys Asn Ala Gly Pro
440 445 450
Glu Asp Ile Gln Ser Arg Ile Gly His Cys Met Leu Phe His Ser
455 460 465
Lys Asn Arg Cys Leu Tyr Val Phe Gly Gly Gln Arg Ser Lys Thr
470 475 480
Tyr Leu Asn Asp Phe Phe Ser Tyr Asp Val Asp Ser Asp His Val
485 490 495
Asp Ile Ile Ser Asp Gly Thr Lys Lys Asp Ser Gly Met Val Pro
500 505 510
Met Thr Gly Phe Thr Gln Arg Ala Thr Ile Asp Pro Glu Leu Asn
515 520 525
Glu Ile His Val Leu Ser Gly Leu Ser Lys Asp Lys Glu Lys Arg
530 535 540
Glu Glu Asn Val Arg Asn Ser Phe Trp Ile Tyr Asp Ile Val Arg
545 550 555
Asn Ser Trp Ser Cys Val Tyr Lys Asn Asp Gln Ala Ala Lys Asp
560 565 570
Asn Pro Thr Lys Ser Leu Gln Glu Glu Glu Pro Cys Pro Arg Phe
575 580 585
Ala His Gln Leu Val Tyr Asp Glu Leu His Lys Val His Tyr Leu
590 595 600
Phe Gly Gly Asn Pro Gly Lys Ser Cys Ser Pro Lys Met Arg Leu
605 610 615
Asp Asp Phe Trp Ser Leu Lys Leu Cys Arg Pro Ser Lys Asp Tyr
620 625 630
Leu Leu Arg His Cys Lys Tyr Leu Ile Arg Lys His Arg Phe Glu
635 640 645
Glu Lys Ala Gln Val Asp Pro Leu Ser Ala Leu Lys Tyr Leu Gln

```
                    650                      655                        660
Asn Asp Leu Tyr Ile Thr Val Asp His Ser Asp Pro Glu Glu Thr
                    665                      670                        675
Lys Glu Phe Gln Leu Leu Ala Ser Ala Leu Phe Lys Ser Gly Ser
                    680                      685                        690
Asp Phe Thr Ala Leu Gly Phe Ser Asp Val Asp His Thr Tyr Ala
                    695                      700                        705
Gln Arg Thr Gln Leu Phe Asp Thr Leu Val Asn Phe Phe Pro Asp
                    710                      715                        720
Ser Met Thr Pro Pro Lys Gly Asn Leu Val Asp Leu Ile Thr Leu
                    725                      730                        735
<210> 5
<211> 424
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1598937CD1

<400> 5
Met Ala Pro Glu Glu Asp Ala Gly Gly Glu Ala Leu Gly Gly Ser
  1               5                   10                  15
Phe Trp Glu Ala Gly Asn Tyr Arg Arg Thr Val Gln Arg Val Glu
                 20                   25                  30
Asp Gly His Arg Leu Cys Gly Asp Leu Val Ser Cys Phe Gln Glu
                 35                   40                  45
Arg Ala Arg Ile Glu Lys Ala Tyr Ala Gln Gln Leu Ala Asp Trp
                 50                   55                  60
Ala Arg Lys Trp Arg Gly Thr Val Glu Lys Gly Pro Gln Tyr Gly
                 65                   70                  75
Thr Leu Glu Lys Ala Trp His Ala Phe Phe Thr Ala Ala Glu Arg
                 80                   85                  90
Leu Ser Ala Leu His Leu Glu Val Arg Glu Lys Leu Gln Gly Gln
                 95                  100                 105
Asp Ser Glu Arg Val Arg Ala Trp Gln Arg Gly Ala Phe His Arg
                110                  115                 120
Pro Val Leu Gly Gly Phe Arg Glu Ser Arg Ala Ala Glu Asp Gly
                125                  130                 135
Phe Arg Lys Ala Gln Lys Pro Trp Leu Lys Arg Leu Lys Glu Val
                140                  145                 150
Glu Ala Ser Lys Lys Ser Tyr His Ala Ala Arg Lys Asp Glu Lys
                155                  160                 165
Thr Ala Gln Thr Arg Glu Ser His Ala Lys Ala Asp Ser Ala Val
                170                  175                 180
Ser Gln Glu Gln Leu Arg Lys Leu Gln Glu Arg Val Glu Arg Cys
                185                  190                 195
Ala Lys Glu Ala Glu Lys Thr Lys Ala Gln Tyr Glu Gln Thr Leu
                200                  205                 210
Ala Glu Leu His Arg Tyr Thr Pro Arg Tyr Met Glu Asp Met Glu
                215                  220                 225
Gln Ala Phe Glu Thr Cys Gln Ala Ala Glu Arg Gln Arg Leu Leu
                230                  235                 240
Phe Phe Lys Asp Met Leu Leu Thr Leu His Gln His Leu Asp Leu
                245                  250                 255
Ser Ser Ser Glu Lys Phe His Glu Leu His Arg Asp Leu His Gln
                260                  265                 270
Gly Ile Glu Ala Ala Ser Asp Glu Glu Asp Leu Arg Trp Trp Arg
                275                  280                 285
Ser Thr His Gly Pro Gly Met Ala Met Asn Trp Pro Gln Phe Glu
                290                  295                 300
Glu Trp Ser Leu Asp Thr Gln Arg Thr Ile Ser Arg Lys Glu Lys
                305                  310                 315
Gly Gly Arg Ser Pro Asp Glu Val Thr Leu Thr Ser Ile Val Pro
                320                  325                 330
Thr Arg Asp Gly Thr Ala Pro Pro Pro Gln Ser Pro Gly Ser Pro
                335                  340                 345
```

Gly Thr Gly Gln Asp Glu Glu Trp Ser Asp Glu Glu Ser Pro Arg
350 355 360
Lys Ala Ala Thr Gly Val Arg Val Arg Ala Leu Tyr Asp Tyr Ala
365 370 375
Gly Gln Glu Ala Asp Glu Leu Ser Phe Arg Ala Gly Glu Glu Leu
380 385 390
Leu Lys Met Ser Glu Glu Asp Glu Gln Gly Trp Cys Gln Gly Gln
395 400 405
Leu Gln Ser Gly Arg Ile Gly Leu Tyr Pro Ala Asn Tyr Val Glu
410 415 420
Cys Val Gly Ala

<210> 6
<211> 420
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1725801CD1

<400> 6
Met Ala Pro Trp Pro Pro Lys Gly Leu Val Pro Ala Val Leu Trp
1 5 10 15
Gly Leu Ser Leu Phe Leu Asn Leu Pro Gly Pro Ile Trp Leu Gln
20 25 30
Pro Ser Pro Pro Pro Gln Ser Ser Pro Pro Pro Gln Pro His Pro
35 40 45
Cys His Thr Cys Arg Gly Leu Val Asp Ser Phe Asn Lys Gly Leu
50 55 60
Glu Arg Thr Ile Arg Asp Asn Phe Gly Gly Gly Asn Thr Ala Trp
65 70 75
Glu Glu Glu Asn Leu Ser Lys Tyr Lys Asp Ser Glu Thr Arg Leu
80 85 90
Val Glu Val Leu Glu Gly Val Cys Ser Lys Ser Asp Phe Glu Cys
95 100 105
His Arg Leu Leu Glu Leu Ser Glu Glu Leu Val Glu Ser Trp Trp
110 115 120
Phe His Lys Gln Gln Glu Ala Pro Asp Leu Phe Gln Trp Leu Cys
125 130 135
Ser Asp Ser Leu Lys Leu Cys Cys Pro Ala Gly Thr Phe Gly Pro
140 145 150
Ser Cys Leu Pro Cys Pro Gly Gly Thr Glu Arg Pro Cys Gly Gly
155 160 165
Tyr Gly Gln Cys Glu Gly Glu Gly Thr Arg Gly Gly Ser Gly His
170 175 180
Cys Asp Cys Gln Ala Gly Tyr Gly Gly Glu Ala Cys Gly Gln Cys
185 190 195
Gly Leu Gly Tyr Phe Glu Ala Glu Arg Asn Ala Ser His Leu Val
200 205 210
Cys Ser Ala Cys Phe Gly Pro Cys Ala Arg Cys Ser Gly Pro Glu
215 220 225
Glu Ser Asn Cys Leu Gln Cys Lys Lys Gly Trp Ala Leu His His
230 235 240
Leu Lys Cys Val Asp Ile Asp Glu Cys Gly Thr Glu Gly Ala Asn
245 250 255
Cys Gly Ala Asp Gln Phe Cys Val Asn Thr Glu Gly Ser Tyr Glu
260 265 270
Cys Arg Asp Cys Ala Lys Ala Cys Leu Gly Cys Met Gly Ala Gly
275 280 285
Pro Gly Arg Cys Lys Lys Cys Ser Pro Gly Tyr Gln Gln Val Gly
290 295 300
Ser Lys Cys Leu Asp Val Asp Glu Cys Glu Thr Glu Val Cys Pro
305 310 315
Gly Glu Asn Lys Gln Cys Glu Asn Thr Glu Gly Gly Tyr Arg Cys
320 325 330
Ile Cys Ala Glu Gly Tyr Lys Gln Met Glu Gly Ile Cys Val Lys

```
                    335                                 340                         345
Glu Gln Ile Pro Glu Ser Ala Gly Phe Phe Ser Glu Met Thr Glu
                    350                                 355                         360
Asp Glu Leu Val Val Leu Gln Gln Met Phe Phe Gly Ile Ile Ile
                    365                                 370                         375
Cys Ala Leu Ala Thr Leu Ala Ala Lys Gly Asp Leu Val Phe Thr
                    380                                 385                         390
Ala Ile Phe Ile Gly Ala Val Ala Ala Met Thr Gly Tyr Trp Leu
                    395                                 400                         405
Ser Glu Arg Ser Asp Arg Val Leu Glu Gly Phe Ile Lys Gly Arg
                    410                                 415                         420
```

<210> 7
<211> 795
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1730482CD1

<400> 7

```
Met Glu Lys Thr Gln Ser Leu Pro Thr Arg Pro Pro Thr Phe Pro
  1               5                  10                  15
Pro Thr Ile Pro Pro Ala Lys Glu Val Cys Lys Ala Ala Lys Ala
                 20                  25                  30
Asp Leu Val Phe Met Val Asp Gly Ser Trp Ser Ile Gly Asp Glu
                 35                  40                  45
Asn Phe Asn Lys Ile Ile Ser Phe Leu Tyr Ser Thr Val Gly Ala
                 50                  55                  60
Leu Asn Lys Ile Gly Thr Asp Gly Thr Gln Val Ala Met Val Gln
                 65                  70                  75
Phe Thr Asp Asp Pro Arg Thr Glu Phe Lys Leu Asn Ala Tyr Lys
                 80                  85                  90
Thr Lys Glu Thr Leu Leu Asp Ala Ile Lys His Ile Ser Tyr Lys
                 95                 100                 105
Gly Gly Asn Thr Lys Thr Gly Lys Ala Ile Lys Tyr Val Arg Asp
                110                 115                 120
Thr Leu Phe Thr Ala Glu Ser Gly Thr Arg Arg Gly Ile Pro Lys
                125                 130                 135
Val Ile Val Val Ile Thr Asp Gly Arg Ser Gln Asp Asp Val Asn
                140                 145                 150
Lys Ile Ser Arg Glu Met Gln Leu Asp Gly Tyr Ser Ile Phe Ala
                155                 160                 165
Ile Gly Val Ala Asp Ala Asp Tyr Ser Glu Leu Val Ser Ile Gly
                170                 175                 180
Ser Lys Pro Ser Ala Arg His Val Phe Phe Val Asp Asp Phe Asp
                185                 190                 195
Ala Phe Lys Lys Ile Glu Asp Glu Leu Ile Thr Phe Val Cys Glu
                200                 205                 210
Thr Ala Ser Ala Thr Cys Pro Val Val His Lys Asp Gly Ile Asp
                215                 220                 225
Leu Ala Gly Phe Lys Met Met Glu Met Phe Gly Leu Val Glu Lys
                230                 235                 240
Asp Phe Ser Ser Val Glu Gly Val Ser Met Glu Pro Gly Thr Phe
                245                 250                 255
Asn Val Phe Pro Cys Tyr Gln Leu His Lys Asp Ala Leu Val Ser
                260                 265                 270
Gln Pro Thr Arg Tyr Leu His Pro Glu Gly Leu Pro Ser Asp Tyr
                275                 280                 285
Thr Ile Ser Phe Leu Phe Arg Ile Leu Pro Asp Thr Pro Gln Glu
                290                 295                 300
Pro Phe Ala Leu Trp Glu Ile Leu Asn Lys Asn Ser Asp Pro Leu
                305                 310                 315
Val Gly Val Ile Leu Asp Asn Gly Gly Lys Thr Leu Thr Tyr Phe
                320                 325                 330
Asn Tyr Asp Gln Ser Gly Asp Phe Gln Thr Val Thr Phe Glu Gly
                335                 340                 345
```

```
Pro Glu Ile Arg Lys Ile Phe Tyr Gly Ser Phe His Lys Leu His
                350                 355                 360
Ile Val Val Ser Glu Thr Leu Val Lys Val Val Ile Asp Cys Lys
                365                 370                 375
Gln Val Gly Glu Lys Ala Met Asn Ala Ser Ala Asn Ile Thr Ser
                380                 385                 390
Asp Gly Val Glu Val Leu Gly Lys Met Val Arg Ser Arg Gly Pro
                395                 400                 405
Gly Gly Asn Ser Ala Pro Phe Gln Leu Gln Met Phe Asp Ile Val
                410                 415                 420
Cys Ser Thr Ser Trp Ala Asn Thr Asp Lys Cys Cys Glu Leu Pro
                425                 430                 435
Gly Leu Arg Asp Asp Glu Ser Cys Pro Asp Leu Pro His Ser Cys
                440                 445                 450
Ser Cys Ser Glu Thr Asn Glu Val Ala Leu Gly Pro Ala Gly Pro
                455                 460                 465
Pro Gly Gly Pro Gly Leu Arg Gly Pro Lys Gly Gln Gln Gly Glu
                470                 475                 480
Pro Gly Pro Lys Gly Pro Asp Gly Pro Arg Gly Glu Ile Gly Leu
                485                 490                 495
Pro Gly Pro Gln Gly Pro Pro Gly Pro Gln Gly Pro Ser Gly Leu
                500                 505                 510
Ser Ile Gln Gly Met Pro Gly Met Pro Gly Glu Lys Gly Glu Lys
                515                 520                 525
Gly Asp Thr Gly Leu Pro Gly Pro Gln Gly Ile Pro Gly Gly Val
                530                 535                 540
Gly Ser Pro Gly Arg Asp Gly Ser Pro Gly Gln Arg Gly Leu Pro
                545                 550                 555
Gly Lys Asp Gly Ser Ser Gly Pro Pro Gly Pro Pro Gly Pro Ile
                560                 565                 570
Gly Ile Pro Gly Thr Pro Gly Val Pro Gly Ile Thr Gly Ser Met
                575                 580                 585
Gly Pro Gln Gly Ala Leu Gly Pro Pro Gly Val Pro Gly Ala Lys
                590                 595                 600
Gly Glu Arg Gly Glu Arg Gly Asp Leu Gln Ser Gln Ala Met Val
                605                 610                 615
Arg Ser Val Ala Arg Gln Val Cys Glu Gln Leu Ile Gln Ser His
                620                 625                 630
Met Ala Arg Tyr Thr Ala Ile Leu Asn Gln Ile Pro Ser His Ser
                635                 640                 645
Ser Ser Ile Arg Thr Val Gln Gly Pro Pro Gly Glu Pro Gly Arg
                650                 655                 660
Pro Gly Ser Pro Gly Ala Pro Gly Glu Gln Gly Pro Pro Gly Thr
                665                 670                 675
Pro Gly Phe Pro Gly Asn Ala Gly Val Pro Gly Thr Pro Gly Glu
                680                 685                 690
Arg Gly Leu Thr Gly Ile Lys Gly Glu Lys Gly Asn Pro Gly Val
                695                 700                 705
Gly Thr Gln Gly Pro Arg Gly Pro Pro Gly Pro Ala Gly Pro Ser
                710                 715                 720
Gly Glu Ser Arg Pro Gly Ser Pro Gly Pro Pro Gly Ser Pro Gly
                725                 730                 735
Pro Arg Gly Pro Pro Gly His Leu Gly Val Pro Gly Pro Gln Gly
                740                 745                 750
Pro Ser Gly Gln Pro Gly Tyr Cys Asp Pro Ser Ser Cys Ser Ala
                755                 760                 765
Tyr Gly Val Arg Ala Pro His Pro Asp Gln Pro Glu Phe Thr Pro
                770                 775                 780
Val Gln Asp Glu Leu Glu Ala Met Glu Leu Trp Gly Pro Gly Val
                785                 790                 795
```

<210> 8
<211> 306
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 1810058CD1

<400> 8
Met Arg Ile Trp Trp Leu Leu Leu Ala Ile Glu Ile Cys Thr Gly
1 5 10 15
Asn Ile Asn Ser Gln Asp Thr Cys Arg Gln Gly His Pro Gly Ile
20 25 30
Pro Gly Asn Pro Gly His Asn Gly Leu Pro Gly Arg Asp Gly Arg
35 40 45
Asp Gly Ala Lys Gly Asp Lys Gly Asp Ala Gly Glu Pro Gly Arg
50 55 60
Pro Gly Ser Pro Gly Lys Asp Gly Thr Ser Gly Glu Lys Gly Glu
65 70 75
Arg Gly Ala Asp Gly Lys Val Glu Ala Lys Gly Ile Lys Gly Asp
80 85 90
Gln Gly Ser Arg Gly Ser Pro Gly Lys His Gly Pro Lys Gly Leu
95 100 105
Ala Gly Pro Met Gly Glu Lys Gly Leu Arg Gly Glu Thr Gly Pro
110 115 120
Gln Gly Gln Lys Gly Asn Lys Gly Asp Val Gly Pro Thr Gly Pro
125 130 135
Glu Gly Pro Arg Gly Asn Ile Gly Pro Leu Gly Pro Thr Gly Leu
140 145 150
Pro Gly Pro Met Gly Pro Ile Gly Lys Pro Gly Pro Lys Gly Glu
155 160 165
Ala Gly Pro Thr Gly Pro Gln Gly Glu Pro Gly Val Arg Gly Ile
170 175 180
Arg Gly Trp Lys Gly Asp Arg Gly Glu Lys Gly Lys Ile Gly Glu
185 190 195
Thr Leu Val Leu Pro Lys Ser Ala Phe Thr Val Gly Leu Thr Val
200 205 210
Leu Ser Lys Phe Pro Ser Ser Asp Val Pro Ile Lys Phe Asp Lys
215 220 225
Ile His Ile Thr Val Phe Ser Arg Asn Val Gln Val Ser Leu Val
230 235 240
Lys Asn Gly Val Lys Ile Leu His Thr Arg Asp Ala Tyr Val Ser
245 250 255
Ser Glu Asp Gln Ala Ser Gly Ser Ile Val Leu Gln Leu Lys Leu
260 265 270
Gly Asp Glu Met Trp Leu Gln Val Thr Gly Gly Glu Arg Phe Asn
275 280 285
Gly Leu Phe Ala Asp Glu Asp Asp Asp Thr Thr Phe Thr Gly Phe
290 295 300
Leu Leu Phe Ser Ser Gln
305

<210> 9
<211> 338
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2040679CD1

<400> 9
Met Tyr Val Leu Ser Pro Val Glu Phe Ile Ile Leu Gln Leu Leu
1 5 10 15
Phe Ile Gln Ala Ile Ser Ser Ser Leu Lys Gly Phe Leu Ser Ala
20 25 30
Met Arg Leu Ala His Arg Gly Cys Asn Val Asp Thr Pro Val Ser
35 40 45
Thr Leu Thr Pro Val Lys Thr Ser Glu Phe Glu Asn Phe Lys Thr
50 55 60
Lys Met Val Ile Thr Ser Lys Lys Asp Tyr Pro Leu Ser Lys Asn
65 70 75
Phe Pro Tyr Ser Leu Glu His Leu Gln Thr Ser Tyr Cys Gly Leu
80 85 90

Val Arg Val Asp Met Arg Met Leu Cys Leu Lys Ser Leu Arg Lys
95 100 105
Leu Asp Leu Ser His Asn His Ile Lys Lys Leu Pro Ala Thr Ile
110 115 120
Gly Asp Leu Ile His Leu Gln Glu Leu Asn Leu Asn Asp Asn His
125 130 135
Leu Glu Ser Phe Ser Val Ala Leu Cys His Ser Thr Leu Gln Lys
140 145 150
Ser Leu Arg Ser Leu Asp Leu Ser Lys Asn Lys Ile Lys Ala Leu
155 160 165
Pro Val Gln Phe Cys Gln Leu Gln Glu Leu Lys Asn Leu Lys Leu
170 175 180
Asp Asp Asn Glu Leu Ile Gln Phe Pro Cys Lys Ile Gly Gln Leu
185 190 195
Ile Asn Leu Arg Phe Leu Ser Ala Ala Arg Asn Lys Leu Pro Phe
200 205 210
Leu Pro Ser Glu Phe Arg Asn Leu Ser Leu Glu Tyr Leu Asp Leu
215 220 225
Phe Gly Asn Thr Phe Glu Gln Pro Lys Val Leu Pro Val Ile Lys
230 235 240
Leu Gln Ala Pro Leu Thr Leu Leu Glu Ser Ser Ala Arg Thr Ile
245 250 255
Leu His Asn Arg Ile Pro Tyr Gly Ser His Ile Ile Pro Phe His
260 265 270
Leu Cys Gln Asp Leu Asp Thr Ala Lys Ile Cys Val Cys Gly Arg
275 280 285
Phe Cys Leu Asn Ser Phe Ile Gln Gly Thr Thr Thr Met Asn Leu
290 295 300
His Ser Val Ala His Thr Val Val Leu Val Asp Asn Leu Gly Gly
305 310 315
Thr Glu Ala Pro Ile Ile Ser Tyr Phe Cys Ser Leu Gly Cys Tyr
320 325 330
Val Asn Ser Ser Asp Met Leu Lys
335

<210> 10
<211> 164
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2960051CD1

<400> 10
Met Lys Ile Ala Val Leu Phe Cys Phe Phe Leu Leu Ile Ile Phe
1 5 10 15
Gln Thr Asp Phe Gly Lys Asn Glu Glu Ile Pro Arg Lys Gln Arg
20 25 30
Arg Lys Ile Tyr His Arg Arg Leu Arg Lys Ser Ser Thr Ser His
35 40 45
Lys His Arg Ser Asn Arg Gln Leu Gly Ile Pro Gln Thr Thr Val
50 55 60
Phe Thr Pro Val Ala Arg Leu Pro Ile Val Asn Phe Asp Tyr Ser
65 70 75
Met Glu Glu Lys Phe Glu Ser Phe Ser Ser Phe Pro Gly Val Glu
80 85 90
Ser Ser Tyr Asn Val Leu Pro Gly Lys Lys Gly His Cys Leu Val
95 100 105
Lys Gly Ile Thr Met Tyr Asn Lys Ala Val Trp Ser Pro Glu Pro
110 115 120
Cys Thr Thr Cys Leu Cys Ser Asp Gly Arg Val Leu Cys Asp Glu
125 130 135
Thr Met Cys His Pro Gln Arg Cys Pro Gln Thr Val Ile Pro Glu
140 145 150
Gly Glu Cys Cys Pro Val Cys Ser Ala Thr Gly Thr Glu Ile
155 160

<210> 11

<211> 327
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3117318CD1

<400> 11
Met Arg Ala Leu Pro Gly Leu Leu Glu Ala Arg Ala Arg Thr Pro
1 5 10 15
Arg Leu Leu Leu Leu Gln Cys Leu Leu Ala Ala Ala Arg Pro Ser
20 25 30
Ser Ala Asp Gly Ser Ala Pro Asp Ser Ala Phe Thr Ser Pro Pro
35 40 45
Leu Arg Glu Glu Ile Met Ala Asn Asn Phe Ser Leu Glu Ser His
50 55 60
Asn Ile Ser Leu Thr Glu His Ser Ser Met Pro Val Glu Lys Asn
65 70 75
Ile Thr Leu Glu Arg Pro Ser Asn Val Asn Leu Thr Cys Gln Phe
80 85 90
Thr Thr Ser Gly Asp Leu Asn Ala Val Asn Val Thr Trp Lys Lys
95 100 105
Asp Gly Glu Gln Leu Glu Asn Asn Tyr Leu Val Ser Ala Thr Gly
110 115 120
Ser Thr Leu Tyr Thr Gln Tyr Arg Phe Thr Ile Ile Asn Ser Lys
125 130 135
Gln Met Gly Ser Tyr Ser Cys Phe Phe Arg Glu Glu Lys Glu Gln
140 145 150
Arg Gly Thr Phe Asn Phe Lys Val Pro Glu Leu His Gly Lys Asn
155 160 165
Lys Pro Leu Ile Ser Tyr Val Gly Asp Ser Thr Val Leu Thr Cys
170 175 180
Lys Cys Gln Asn Cys Phe Pro Leu Asn Trp Thr Trp Tyr Ser Ser
185 190 195
Asn Gly Ser Val Lys Val Pro Val Gly Val Gln Met Asn Lys Tyr
200 205 210
Val Ile Asn Gly Thr Tyr Ala Asn Glu Thr Lys Leu Lys Ile Thr
215 220 225
Gln Leu Leu Glu Glu Asp Gly Glu Ser Tyr Trp Cys Arg Ala Leu
230 235 240
Phe Gln Leu Gly Glu Ser Glu Glu His Ile Glu Leu Val Val Leu
245 250 255
Ser Tyr Leu Val Pro Leu Lys Pro Phe Leu Val Ile Val Ala Glu
260 265 270
Val Ile Leu Leu Val Ala Thr Ile Leu Leu Cys Glu Lys Tyr Thr
275 280 285
Gln Lys Lys Lys Lys His Ser Asp Glu Gly Lys Glu Phe Glu Gln
290 295 300
Ile Glu Gln Leu Lys Ser Asp Asp Ser Asn Gly Ile Glu Asn Asn
305 310 315
Val Pro Arg His Arg Lys Asn Glu Ser Leu Gly Gln
320 325

<210> 12
<211> 716
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3486992CD1

<400> 12
Met Ala Arg Met Ser Phe Val Ile Ala Ala Cys Gln Leu Val Leu
1 5 10 15
Gly Leu Leu Met Thr Ser Leu Thr Glu Ser Ser Ile Gln Asn Ser
20 25 30

Glu Cys Pro Gln Leu Cys Val Cys Glu Ile Arg Pro Trp Phe Thr
35 40 45
Pro Gln Ser Thr Tyr Arg Glu Ala Thr Thr Val Asp Cys Asn Asp
50 55 60
Leu Arg Leu Thr Arg Ile Pro Ser Asn Leu Ser Ser Asp Thr Gln
65 70 75
Val Leu Leu Leu Gln Ser Asn Asn Ile Ala Lys Thr Val Asp Glu
80 85 90
Leu Gln Gln Leu Phe Asn Leu Thr Glu Leu Asp Phe Ser Gln Asn
95 100 105
Asn Phe Thr Asn Ile Lys Glu Val Gly Leu Ala Asn Leu Thr Gln
110 115 120
Leu Thr Thr Leu His Leu Glu Glu Asn Gln Ile Thr Glu Met Thr
125 130 135
Asp Tyr Cys Leu Gln Asp Leu Ser Asn Leu Gln Glu Leu Tyr Ile
140 145 150
Asn His Asn Gln Ile Ser Thr Ile Ser Ala His Ala Phe Ala Gly
155 160 165
Leu Lys Asn Leu Leu Arg Leu His Leu Asn Ser Asn Lys Leu Lys
170 175 180
Val Ile Asp Ser Arg Trp Phe Asp Ser Thr Pro Asn Leu Glu Ile
185 190 195
Leu Met Ile Gly Glu Asn Pro Val Ile Gly Ile Leu Asp Met Asn
200 205 210
Phe Lys Pro Leu Ala Asn Leu Arg Ser Leu Val Leu Ala Gly Met
215 220 225
Tyr Leu Thr Asp Ile Pro Gly Asn Ala Leu Val Gly Leu Asp Ser
230 235 240
Leu Glu Ser Leu Ser Phe Tyr Asp Asn Lys Leu Val Lys Val Pro
245 250 255
Gln Leu Ala Leu Gln Lys Val Pro Asn Leu Lys Phe Leu Asp Leu
260 265 270
Asn Lys Asn Pro Ile His Lys Ile Gln Glu Gly Asp Phe Lys Asn
275 280 285
Met Leu Arg Leu Lys Glu Leu Gly Ile Asn Asn Met Gly Glu Leu
290 295 300
Val Ser Val Asp Arg Tyr Ala Leu Asp Asn Leu Pro Glu Leu Thr
305 310 315
Lys Leu Glu Ala Thr Asn Asn Pro Lys Leu Ser Tyr Ile His Arg
320 325 330
Leu Ala Phe Arg Ser Val Pro Ala Leu Glu Ser Leu Met Leu Asn
335 340 345
Asn Asn Ala Leu Asn Ala Ile Tyr Gln Lys Thr Val Glu Ser Leu
350 355 360
Pro Asn Leu Arg Glu Ile Ser Ile His Ser Asn Pro Leu Arg Cys
365 370 375
Asp Cys Val Ile His Trp Ile Asn Ser Asn Lys Thr Asn Ile Arg
380 385 390
Phe Met Glu Pro Leu Ser Met Phe Cys Ala Met Pro Pro Glu Tyr
395 400 405
Lys Gly His Gln Val Lys Glu Val Leu Ile Gln Asp Ser Ser Glu
410 415 420
Gln Cys Leu Pro Met Ile Ser His Asp Ser Phe Pro Asn Arg Leu
425 430 435
Asn Val Asp Ile Gly Thr Thr Val Phe Leu Asp Cys Arg Ala Met
440 445 450
Ala Glu Pro Glu Pro Glu Ile Tyr Trp Val Thr Pro Ile Gly Asn
455 460 465
Lys Ile Thr Val Glu Thr Leu Ser Asp Lys Tyr Lys Leu Ser Ser
470 475 480
Glu Gly Thr Leu Glu Ile Ser Asn Ile Gln Ile Glu Asp Ser Gly
485 490 495
Arg Tyr Thr Cys Val Ala Gln Asn Val Gln Gly Ala Asp Thr Arg
500 505 510
Val Ala Thr Ile Lys Val Asn Gly Thr Leu Leu Asp Gly Thr Gln
515 520 525
Val Leu Lys Ile Tyr Val Lys Gln Thr Glu Ser His Ser Ile Leu

530 535 540
Val Ser Trp Lys Val Asn Ser Asn Val Met Thr Ser Asn Leu Lys
545 550 555
Trp Ser Ser Ala Thr Met Lys Ile Asp Asn Pro His Ile Thr Tyr
560 565 570
Thr Ala Arg Val Pro Val Asp Val His Glu Tyr Asn Leu Thr His
575 580 585
Leu Gln Pro Ser Thr Asp Tyr Glu Val Cys Leu Thr Val Ser Asn
590 595 600
Ile His Gln Gln Thr Gln Lys Ser Cys Val Asn Val Thr Thr Lys
605 610 615
Asn Ala Ala Phe Ala Val Asp Ile Ser Asp Gln Glu Thr Ser Thr
620 625 630
Ala Leu Ala Ala Val Met Gly Ser Met Phe Ala Val Ile Ser Leu
635 640 645
Ala Ser Ile Ala Val Tyr Phe Ala Lys Arg Phe Lys Arg Lys Asn
650 655 660
Tyr His His Ser Leu Lys Lys Tyr Met Gln Lys Thr Ser Ser Ile
665 670 675
Pro Leu Asn Glu Leu Tyr Pro Pro Leu Ile Asn Leu Trp Glu Gly
680 685 690
Asp Ser Glu Lys Asp Lys Asp Gly Ser Ala Asp Thr Lys Pro Thr
695 700 705
Gln Val Asp Thr Ser Arg Ser Tyr Tyr Met Trp
710 715

<210> 13
<211> 665
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4568384CD1

<400> 13
Met Val Leu Val Phe His Lys Gly Glu Leu Gly His Pro Leu Glu
1 5 10 15
Gln Ser Thr Asp Trp Pro Lys Ser Pro Lys Thr Pro Thr Gly Leu
20 25 30
Arg Arg Gly Arg Gln Cys Ile Arg Pro Ala Glu Ile Val Ala Ser
35 40 45
Leu Leu Glu Gly Glu Glu Asn Thr Cys Gly Lys Gln Lys Pro Lys
50 55 60
Glu Asn Asn Leu Lys Pro Lys Phe Gln Ala Phe Lys Gly Val Gly
65 70 75
Cys Leu Tyr Glu Lys Glu Ser Met Lys Lys Ser Leu Lys Asp Ser
80 85 90
Val Ala Ser Asn Asn Lys Asp Gln Asn Ser Met Lys His Glu Asp
95 100 105
Pro Ser Ile Ile Ser Met Glu Asp Gly Ser Pro Tyr Val Asn Gly
110 115 120
Ser Leu Gly Glu Val Thr Pro Cys Gln His Ala Lys Lys Ala Asn
125 130 135
Gly Pro Asn Tyr Ile Gln Pro Gln Lys Arg Gln Thr Thr Phe Glu
140 145 150
Ser Gln Asp Arg Lys Ala Val Ser Pro Ser Ser Ser Glu Lys Arg
155 160 165
Ser Lys Asn Pro Ile Ser Arg Pro Leu Glu Gly Lys Lys Ser Leu
170 175 180
Ser Leu Ser Ala Lys Thr His Asn Ile Gly Phe Asp Lys Asp Ser
185 190 195
Cys His Ser Thr Thr Lys Thr Glu Ala Ser Gln Glu Glu Arg Ser
200 205 210
Asp Ser Ser Gly Leu Thr Ser Leu Lys Lys Ser Pro Lys Val Ser
215 220 225
Ser Lys Asp Thr Arg Glu Ile Lys Thr Asp Phe Ser Leu Ser Ile
230 235 240

```
Ser Asn Ser Ser Asp Val Ser Ala Lys Asp Lys His Ala Glu Asp
                245                 250                 255
Asn Glu Lys Arg Leu Ala Ala Leu Glu Ala Arg Gln Lys Ala Lys
                260                 265                 270
Glu Val Gln Lys Lys Leu Val His Asn Ala Leu Ala Asn Leu Asp
                275                 280                 285
Gly His Pro Glu Asp Lys Pro Thr His Ile Ile Phe Gly Ser Asp
                290                 295                 300
Ser Glu Cys Glu Thr Glu Glu Thr Ser Thr Gln Glu Gln Ser His
                305                 310                 315
Pro Gly Glu Glu Trp Val Lys Glu Ser Met Gly Lys Thr Ser Gly
                320                 325                 330
Lys Leu Phe Asp Ser Ser Asp Asp Asp Glu Ser Asp Ser Glu Asp
                335                 340                 345
Asp Ser Asn Arg Phe Lys Ile Lys Pro Gln Phe Glu Gly Arg Ala
                350                 355                 360
Gly Gln Lys Leu Met Asp Leu Gln Ser His Phe Gly Thr Asp Asp
                365                 370                 375
Arg Phe Arg Met Asp Ser Arg Phe Leu Glu Thr Asp Ser Glu Glu
                380                 385                 390
Glu Gln Glu Glu Val Asn Glu Lys Lys Thr Ala Glu Glu Glu Glu
                395                 400                 405
Leu Ala Glu Glu Lys Lys Lys Ala Leu Asn Val Val Gln Ser Val
                410                 415                 420
Leu Gln Ile Asn Leu Ser Asn Ser Thr Asn Arg Gly Ser Val Ala
                425                 430                 435
Ala Lys Lys Phe Lys Asp Ile Ile His Tyr Asp Pro Thr Lys Gln
                440                 445                 450
Asp His Ala Thr Tyr Glu Arg Lys Arg Asp Asp Lys Pro Lys Glu
                455                 460                 465
Ser Lys Ala Lys Arg Lys Lys Lys Arg Glu Glu Ala Glu Lys Leu
                470                 475                 480
Pro Glu Val Ser Lys Glu Met Tyr Tyr Asn Ile Ala Met Asp Leu
                485                 490                 495
Lys Glu Ile Phe Gln Thr Thr Lys Tyr Thr Ser Glu Lys Glu Glu
                500                 505                 510
Gly Thr Pro Trp Asn Glu Asp Cys Gly Lys Glu Lys Pro Glu Glu
                515                 520                 525
Ile Gln Asp Pro Ala Ala Leu Thr Ser Asp Ala Glu Gln Pro Ser
                530                 535                 540
Gly Phe Thr Phe Ser Phe Phe Asp Ser Asp Thr Lys Asp Ile Lys
                545                 550                 555
Glu Glu Thr Tyr Arg Val Glu Thr Val Lys Pro Gly Lys Ile Val
                560                 565                 570
Trp Gln Glu Asp Pro Arg Leu Gln Asp Ser Ser Ser Glu Glu Glu
                575                 580                 585
Asp Val Thr Glu Glu Thr Asp His Arg Asn Ser Ser Pro Gly Glu
                590                 595                 600
Ala Ser Leu Leu Glu Lys Glu Thr Thr Arg Phe Phe Phe Phe Ser
                605                 610                 615
Lys Asn Asp Glu Arg Leu Gln Gly Ser Asp Leu Phe Trp Arg Gly
                620                 625                 630
Val Gly Ser Asn Met Ser Arg Asn Ser Trp Glu Ala Arg Thr Thr
                635                 640                 645
Asn Leu Arg Met Asp Cys Arg Lys Lys His Lys Asp Ala Lys Arg
                650                 655                 660
Lys Met Lys Pro Lys
                665
```

<210> 14
<211> 547
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4586187CD1

<400> 14

Met Tyr Ser His Asn Val Val Ile Met Asn Leu Asn Asn Leu Asn
1 5 10 15
Leu Thr Gln Val Gln Gln Arg Asn Leu Ile Thr Asn Leu Gln Arg
20 25 30
Ser Val Asp Asp Thr Ser Gln Ala Ile Gln Arg Ile Lys Asn Asp
35 40 45
Phe Gln Asn Leu Gln Gln Val Phe Leu Gln Ala Lys Lys Asp Thr
50 55 60
Asp Trp Leu Lys Glu Lys Val Gln Ser Leu Gln Thr Leu Ala Ala
65 70 75
Asn Asn Ser Ala Leu Ala Lys Ala Asn Asn Asp Thr Leu Glu Asp
80 85 90
Met Asn Ser Gln Leu Asn Ser Phe Thr Gly Gln Met Glu Asn Ile
95 100 105
Thr Thr Ile Ser Gln Ala Asn Glu Gln Asn Leu Lys Asp Leu Gln
110 115 120
Asp Leu His Lys Asp Ala Glu Asn Arg Thr Ala Ile Lys Phe Asn
125 130 135
Gln Leu Glu Glu Arg Phe Gln Leu Phe Glu Thr Asp Ile Val Asn
140 145 150
Ile Ile Ser Asn Ile Ser Tyr Thr Ala His His Leu Arg Thr Leu
155 160 165
Thr Ser Asn Leu Asn Glu Val Arg Thr Thr Cys Thr Asp Thr Leu
170 175 180
Thr Lys His Thr Asp Asp Leu Thr Ser Leu Asn Asn Thr Leu Ala
185 190 195
Asn Ile Arg Leu Asp Ser Val Ser Leu Arg Met Gln Gln Asp Leu
200 205 210
Met Arg Ser Arg Leu Asp Thr Glu Val Ala Asn Leu Ser Val Ile
215 220 225
Met Glu Glu Met Lys Leu Val Asp Ser Lys His Gly Gln Leu Ile
230 235 240
Lys Asn Phe Thr Ile Leu Gln Gly Pro Pro Gly Pro Arg Gly Pro
245 250 255
Arg Gly Asp Arg Gly Ser Gln Gly Pro Pro Gly Pro Thr Gly Asn
260 265 270
Lys Gly Gln Lys Gly Glu Lys Gly Glu Pro Gly Pro Pro Gly Pro
275 280 285
Ala Gly Glu Arg Gly Pro Ile Gly Pro Ala Gly Pro Pro Gly Glu
290 295 300
Arg Gly Gly Lys Gly Ser Lys Gly Ser Gln Gly Pro Lys Gly Ser
305 310 315
Arg Gly Ser Pro Gly Lys Pro Gly Pro Gln Gly Pro Ser Gly Asp
320 325 330
Pro Gly Pro Pro Gly Pro Pro Gly Lys Glu Gly Leu Pro Gly Pro
335 340 345
Gln Gly Pro Pro Gly Phe Gln Gly Leu Gln Gly Thr Val Gly Glu
350 355 360
Pro Gly Val Pro Gly Pro Arg Gly Leu Pro Gly Leu Pro Gly Val
365 370 375
Pro Gly Met Pro Gly Pro Lys Gly Pro Pro Gly Pro Pro Gly Pro
380 385 390
Ser Gly Ala Val Val Pro Leu Ala Leu Gln Asn Glu Pro Thr Pro
395 400 405
Ala Pro Glu Asp Asn Ser Cys Pro Pro His Trp Lys Asn Phe Thr
410 415 420
Asp Lys Cys Tyr Tyr Phe Ser Val Glu Lys Glu Ile Phe Glu Asp
425 430 435
Ala Lys Leu Phe Cys Glu Asp Lys Ser Ser His Leu Val Phe Ile
440 445 450
Asn Thr Arg Glu Glu Gln Gln Trp Ile Lys Lys Gln Met Val Gly
455 460 465
Arg Glu Ser His Trp Ile Gly Leu Thr Asp Ser Glu Arg Glu Asn
470 475 480
Glu Trp Lys Trp Leu Asp Gly Thr Ser Pro Asp Tyr Lys Asn Trp
485 490 495

Lys Ala Gly Gln Pro Asp Asn Trp Gly His Gly His Gly Pro Gly
500 505 510
Glu Asp Cys Ala Gly Leu Ile Tyr Ala Gly Gln Trp Asn Asp Phe
515 520 525
Gln Cys Glu Asp Val Asn Asn Phe Ile Cys Glu Lys Asp Arg Glu
530 535 540
Thr Val Leu Ser Ser Ala Leu
545

<210> 15
<211> 109
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 401801CD1

<400> 15
Met Tyr Phe Asn Leu Gln Glu Asn Ile Phe Met Tyr Gly Gly Arg
1 5 10 15
Ile Glu Thr Asn Asp Gly Asn Val Thr Asp Glu Leu Trp Val Phe
20 25 30
Asn Ile His Ser Gln Ser Trp Ser Thr Lys Thr Pro Thr Val Leu
35 40 45
Gly His Gly Gln Gln Tyr Ala Val Glu Gly His Ser Ala His Ile
50 55 60
Met Glu Leu Asp Ser Arg Asp Val Val Met Ile Ile Ile Phe Gly
65 70 75
Tyr Ser Ala Ile Tyr Gly Tyr Thr Ser Ser Ile Gln Glu Tyr His
80 85 90
Ile Cys Glu Leu Leu Lys Asn Cys Asn Phe Phe Ile Asp Trp Glu
95 100 105
Cys Phe Ser Leu

<210> 16
<211> 192
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1721842CD1

<400> 16
Met Asn Lys Arg Asp Tyr Met Asn Thr Ser Val Gln Glu Pro Pro
1 5 10 15
Leu Asp Tyr Ser Phe Arg Ser Ile His Val Ile Gln Asp Leu Val
20 25 30
Asn Glu Glu Pro Arg Thr Gly Leu Arg Pro Leu Lys Arg Ser Lys
35 40 45
Ser Gly Lys Ser Leu Thr Gln Ser Leu Trp Leu Asn Asn Asn Val
50 55 60
Leu Asn Asp Leu Arg Asp Phe Asn Gln Val Ala Ser Gln Leu Leu
65 70 75
Glu His Pro Glu Asn Leu Ala Trp Ile Asp Leu Ser Phe Asn Asp
80 85 90
Leu Thr Ser Ile Asp Pro Val Leu Thr Thr Phe Phe Asn Leu Ser
95 100 105
Val Leu Tyr Leu His Gly Asn Ser Ile Gln Arg Leu Gly Glu Val
110 115 120
Asn Lys Leu Ala Val Leu Pro Arg Leu Arg Ser Leu Thr Leu His
125 130 135
Gly Asn Pro Met Glu Glu Glu Lys Gly Tyr Arg Gln Tyr Val Leu
140 145 150
Cys Thr Leu Ser Arg Ile Thr Thr Phe Asp Phe Ser Gly Val Thr
155 160 165
Lys Ala Asp Arg Thr Thr Ala Glu Val Trp Lys Arg Met Asn Ile

```
                170                 175                 180
Lys Pro Lys Lys Ala Trp Thr Lys Gln Asn Thr Leu
                185                 190

<210> 17
<211> 575
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1833221CD1

<400> 17
Met Val Leu Gly Ser Phe Gly Thr Asp Leu Met Arg Glu Arg Arg
  1               5                  10                  15
Asp Leu Glu Arg Arg Thr Asp Ser Ser Ile Ser Asn Leu Met Asp
                 20                  25                  30
Tyr Ser His Arg Ser Gly Asp Phe Thr Thr Ser Ser Tyr Val Gln
                 35                  40                  45
Asp Arg Val Pro Ser Tyr Ser Gln Gly Ala Arg Pro Lys Glu Asn
                 50                  55                  60
Ser Met Ser Thr Leu Gln Leu Asn Thr Ser Ser Thr Asn His Gln
                 65                  70                  75
Leu Pro Ser Glu His Gln Thr Ile Leu Ser Ser Arg Asp Ser Arg
                 80                  85                  90
Asn Ser Leu Arg Ser Asn Phe Ser Ser Arg Glu Ser Glu Ser Ser
                 95                 100                 105
Arg Ser Asn Thr Gln Pro Gly Phe Ser Tyr Ser Ser Ser Arg Asp
                110                 115                 120
Glu Ala Pro Ile Ile Ser Asn Ser Glu Arg Val Val Ser Ser Gln
                125                 130                 135
Arg Pro Phe Gln Glu Ser Ser Asp Asn Glu Gly Arg Arg Thr Thr
                140                 145                 150
Arg Arg Leu Leu Ser Arg Ile Ala Ser Ser Met Ser Ser Thr Phe
                155                 160                 165
Phe Ser Arg Arg Ser Ser Gln Asp Ser Leu Asn Thr Arg Ser Leu
                170                 175                 180
Asn Ser Glu Asn Ser Tyr Val Ser Pro Arg Ile Leu Thr Ala Ser
                185                 190                 195
Gln Ser Arg Ser Asn Val Pro Ser Ala Ser Glu Val Pro Asp Asn
                200                 205                 210
Arg Ala Ser Glu Ala Ser Gln Gly Phe Arg Phe Leu Arg Arg Arg
                215                 220                 225
Trp Gly Leu Ser Ser Leu Ser His Asn His Ser Ser Glu Ser Asp
                230                 235                 240
Ser Glu Asn Phe Asn Gln Glu Ser Glu Gly Arg Asn Thr Gly Pro
                245                 250                 255
Trp Leu Ser Ser Ser Leu Arg Asn Arg Cys Thr Pro Leu Phe Ser
                260                 265                 270
Arg Arg Arg Arg Glu Gly Arg Asp Glu Ser Ser Arg Ile Pro Thr
                275                 280                 285
Ser Asp Thr Ser Ser Arg Ser His Ile Phe Arg Arg Glu Ser Asn
                290                 295                 300
Glu Val Val His Leu Glu Ala Gln Asn Asp Pro Leu Gly Ala Ala
                305                 310                 315
Ala Asn Arg Pro Gln Ala Ser Ala Ala Ser Ser Ser Ala Thr Thr
                320                 325                 330
Gly Gly Ser Thr Ser Asp Ser Ala Gln Gly Gly Arg Asn Thr Gly
                335                 340                 345
Ile Ser Gly Ile Leu Pro Gly Ser Leu Phe Arg Phe Ala Val Pro
                350                 355                 360
Pro Ala Leu Gly Ser Asn Leu Thr Asp Asn Val Met Ile Thr Val
                365                 370                 375
Asp Ile Ile Pro Ser Gly Trp Asn Ser Ala Asp Gly Lys Ser Asp
                380                 385                 390
Lys Thr Lys Ser Ala Pro Ser Arg Asp Pro Glu Arg Leu Gln Lys
                395                 400                 405
```

```
Ile Lys Glu Ser Leu Leu Leu Glu Asp Ser Glu Glu Glu Glu Gly
                410                 415                 420
Asp Leu Cys Arg Ile Cys Gln Met Ala Ala Ala Ser Ser Ser Asn
                425                 430                 435
Leu Leu Ile Glu Pro Cys Lys Cys Thr Gly Ser Leu Gln Tyr Val
                440                 445                 450
His Gln Asp Cys Met Lys Lys Trp Leu Gln Ala Lys Ile Asn Ser
                455                 460                 465
Gly Ser Ser Leu Glu Ala Val Thr Thr Cys Glu Leu Cys Lys Glu
                470                 475                 480
Lys Leu Glu Leu Asn Leu Glu Asp Phe Asp Ile His Glu Leu His
                485                 490                 495
Arg Ala His Ala Asn Glu Gln Ala Glu Tyr Glu Phe Ile Ser Ser
                500                 505                 510
Gly Leu Tyr Leu Val Val Leu Leu His Leu Cys Glu Gln Ser Phe
                515                 520                 525
Ser Asp Met Met Gly Asn Thr Asn Glu Pro Ser Thr Arg Val Arg
                530                 535                 540
Phe Ile Asn Leu Ala Arg Thr Leu Gln Ala His Met Glu Asp Leu
                545                 550                 555
Glu Thr Ser Glu Asp Asp Ser Glu Glu Asp Gly Asp His Asn Arg
                560                 565                 570
Thr Phe Asp Ile Ala
                575

<210> 18
<211> 342
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2041168CD1

<400> 18
Met Ala Glu Gly Gly Ser Gly Asp Val Asp Asp Ala Gly Asp Cys
 1               5                  10                  15
Ser Gly Ala Arg Tyr Asn Asp Trp Ser Asp Asp Asp Asp Asp Ser
                20                  25                  30
Asn Glu Ser Lys Ser Ile Val Trp Tyr Pro Pro Trp Ala Arg Ile
                35                  40                  45
Gly Thr Glu Ala Gly Thr Arg Ala Arg Ala Arg Ala Arg Ala Arg
                50                  55                  60
Ala Thr Arg Ala Arg Arg Ala Val Gln Lys Arg Ala Ser Pro Asn
                65                  70                  75
Ser Asp Asp Thr Val Leu Ser Pro Gln Glu Leu Gln Lys Val Leu
                80                  85                  90
Cys Leu Val Glu Met Ser Glu Lys Pro Tyr Ile Leu Glu Ala Ala
                95                  100                 105
Leu Ile Ala Leu Gly Asn Asn Ala Ala Tyr Ala Phe Asn Arg Asp
                110                 115                 120
Ile Ile Arg Asp Leu Gly Gly Leu Pro Ile Val Ala Lys Ile Leu
                125                 130                 135
Asn Thr Arg Asp Pro Ile Val Lys Glu Lys Ala Leu Ile Val Leu
                140                 145                 150
Asn Asn Leu Ser Val Asn Ala Glu Asn Gln Arg Arg Leu Lys Val
                155                 160                 165
Tyr Met Asn Gln Val Cys Asp Asp Thr Ile Thr Ser Arg Leu Asn
                170                 175                 180
Ser Ser Val Gln Leu Ala Gly Leu Arg Leu Leu Thr Asn Met Thr
                185                 190                 195
Val Thr Asn Glu Tyr Gln His Met Leu Ala Asn Ser Ile Ser Asp
                200                 205                 210
Phe Phe Arg Leu Phe Ser Ala Gly Asn Glu Glu Thr Lys Leu Gln
                215                 220                 225
Val Leu Lys Leu Leu Leu Asn Leu Ala Glu Asn Pro Ala Met Thr
                230                 235                 240
Arg Glu Leu Leu Arg Ala Gln Val Pro Ser Ser Leu Gly Ser Leu
```

```
                        245                                  250                                255
Phe Asn Lys Lys Glu Asn Lys Glu Val Ile Leu Lys Leu Leu Val
                        260                                  265                                270
Ile Phe Glu Asn Ile Asn Asp Asn Phe Lys Trp Glu Glu Asn Glu
                        275                                  280                                285
Pro Thr Gln Asn Gln Phe Gly Glu Gly Ser Leu Phe Phe Phe Leu
                        290                                  295                                300
Lys Glu Phe Gln Val Cys Ala Asp Lys Val Leu Gly Ile Glu Ser
                        305                                  310                                315
His His Asp Phe Leu Val Lys Val Lys Val Gly Lys Phe Met Ala
                        320                                  325                                330
Lys Leu Ala Glu His Met Phe Pro Lys Ser Gln Glu
                        335                                  340
<210> 19
<211> 110
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2365794CD1

<400> 19
Met Ala Ala Val Val Ala Lys Arg Glu Gly Pro Pro Phe Ile Ser
  1               5                   10                  15
Glu Ala Ala Val Arg Gly Asn Ala Ala Val Leu Asp Tyr Cys Arg
                 20                   25                  30
Thr Ser Val Ser Ala Leu Ser Gly Ala Thr Ala Gly Ile Leu Gly
                 35                   40                  45
Leu Thr Gly Leu Tyr Gly Phe Ile Phe Tyr Leu Leu Ala Ser Val
                 50                   55                  60
Leu Leu Ser Leu Leu Leu Ile Leu Lys Ala Gly Arg Arg Trp Asn
                 65                   70                  75
Lys Tyr Phe Lys Ser Arg Arg Pro Leu Phe Thr Gly Gly Leu Ile
                 80                   85                  90
Gly Gly Leu Phe Thr Tyr Val Leu Phe Trp Thr Phe Leu Tyr Gly
                 95                  100                 105
Met Val His Val Tyr
                110
<210> 20
<211> 571
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2618452CD1

<400> 20
Met Pro Thr Gly Thr Ile Pro Pro Pro Thr Thr Leu Lys Ala Thr
  1               5                   10                  15
Gly Ser Thr His Thr Ala Pro Pro Met Met Pro Thr Thr Ser Gly
                 20                   25                  30
Thr Ser Gln Ala Ser Ser Ser Phe Asn Thr Ala Lys Thr Ser Thr
                 35                   40                  45
Ser Leu His Ser His Thr Ser Ser Thr His His Pro Glu Val Thr
                 50                   55                  60
Pro Thr Ser Ile Thr Asn Ile Thr Leu Asn Pro Thr Ser Ile Gly
                 65                   70                  75
Thr Trp Thr Pro Val Ala His Thr Thr Ser Ala Thr Ser Ser Arg
                 80                   85                  90
Leu Thr Thr Pro Phe Thr Thr His Ser Pro Pro Thr Gly Ser Ser
                 95                  100                 105
Pro Ile Ser Ser Thr Gly Pro Met Thr Ala Thr Ser Phe Gln Thr
                110                  115                 120
Thr Thr Tyr Tyr Thr Pro Pro Ser His Pro Gln Thr Thr Leu Pro
                125                  130                 135
```

Thr His Val Pro Pro Phe Ser Thr Ser Leu Val Thr Pro Ser Thr
140 145 150
His Thr Val Ile Ile Thr Thr His Thr Gln Met Ala Thr Ser Ala
155 160 165
Ser Ile His Ser Thr Pro Thr Gly Thr Val Pro Pro Pro Thr Thr
170 175 180
Leu Lys Ala Thr Gly Ser Thr His Thr Ala Pro Pro Met Thr Val
185 190 195
Thr Thr Ser Gly Thr Ser Gln Thr His Ser Ser Phe Ser Thr Ala
200 205 210
Thr Ala Ser Ser Ser Phe Ile Ser Ser Ser Ser Trp Ser Ser Trp
215 220 225
Leu Pro Gln Asn Ser Ser Ser Arg Pro Pro Ser Ser Pro Ile Thr
230 235 240
Thr Gln Leu Pro His Leu Ser Ser Ala Thr Thr Pro Val Ser Thr
245 250 255
Thr Asn Gln Leu Ser Ser Ser Phe Ser Pro Ser Pro Ser Ala Pro
260 265 270
Ser Thr Val Ser Ser Tyr Val Pro Ser Ser His Ser Ser Pro Gln
275 280 285
Thr Ser Ser Pro Ser Val Gly Thr Ser Ser Ser Phe Val Ser Ala
290 295 300
Pro Val His Ser Thr Thr Leu Ser Ser Gly Ser His Ser Ser Leu
305 310 315
Ser Thr His Pro Thr Thr Ala Ser Val Ser Ala Ser Pro Leu Phe
320 325 330
Pro Ser Ser Pro Ala Ala Ser Thr Thr Ile Arg Ala Thr Leu Pro
335 340 345
His Thr Ile Ser Ser Pro Phe Thr Leu Ser Ala Leu Leu Pro Ile
350 355 360
Ser Thr Val Thr Val Ser Pro Thr Pro Ser Ser His Leu Ala Ser
365 370 375
Ser Thr Ile Ala Phe Pro Ser Thr Pro Arg Thr Thr Ala Ser Thr
380 385 390
His Thr Ala Pro Ala Phe Ser Ser Gln Ser Thr Thr Ser Arg Ser
395 400 405
Thr Ser Leu Thr Thr Arg Val Pro Thr Ser Gly Phe Val Ser Leu
410 415 420
Thr Ser Gly Val Thr Gly Ile Pro Thr Ser Pro Val Thr Asn Leu
425 430 435
Thr Thr Arg His Pro Gly Pro Thr Leu Ser Pro Thr Thr Arg Phe
440 445 450
Leu Thr Ser Ser Leu Thr Ala His Gly Ser Thr Pro Ala Ser Ala
455 460 465
Pro Val Ser Ser Leu Gly Thr Pro Thr Pro Thr Ser Pro Gly Val
470 475 480
Cys Ser Val Arg Glu Gln Gln Glu Glu Ile Thr Phe Lys Gly Cys
485 490 495
Met Ala Asn Val Thr Val Thr Arg Cys Glu Gly Ala Cys Ile Ser
500 505 510
Ala Ala Ser Phe Asn Ile Ile Thr Gln Gln Val Asp Ala Arg Cys
515 520 525
Ser Cys Cys Arg Pro Leu His Ser Tyr Glu Gln Gln Leu Glu Leu
530 535 540
Pro Cys Pro Asp Pro Ser Thr Pro Gly Arg Arg Leu Val Leu Thr
545 550 555
Leu Gln Val Phe Ser His Cys Val Cys Ser Ser Val Ala Cys Gly
560 565 570

Asp
<210> 21
<211> 262
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2622288CD1

<400> 21
Met Val Ala Trp Arg Ser Ala Phe Leu Val Cys Leu Ala Phe Ser
1 5 10 15
Leu Ala Thr Leu Val Gln Arg Gly Ser Gly Asp Phe Asp Asp Phe
20 25 30
Asn Leu Glu Asp Ala Val Lys Glu Thr Ser Ser Val Lys Gln Pro
35 40 45
Trp Asp His Thr Thr Thr Thr Thr Thr Asn Arg Pro Gly Thr Thr
50 55 60
Arg Ala Pro Ala Lys Pro Pro Gly Ser Gly Leu Asp Leu Ala Asp
65 70 75
Ala Leu Asp Asp Gln Asp Asp Gly Arg Arg Lys Pro Gly Ile Gly
80 85 90
Gly Arg Glu Arg Trp Asn His Val Thr Thr Thr Thr Lys Arg Pro
95 100 105
Val Thr Thr Arg Ala Pro Ala Asn Thr Leu Gly Asn Asp Phe Asp
110 115 120
Leu Ala Asp Ala Leu Asp Asp Arg Asn Asp Arg Asp Asp Gly Arg
125 130 135
Arg Lys Pro Ile Ala Gly Gly Gly Gly Phe Ser Asp Lys Asp Leu
140 145 150
Glu Asp Ile Val Gly Gly Gly Glu Tyr Lys Pro Asp Lys Gly Lys
155 160 165
Gly Asp Gly Arg Tyr Gly Ser Asn Asp Asp Pro Gly Ser Gly Met
170 175 180
Val Ala Glu Pro Gly Thr Ile Ala Gly Val Ala Ser Ala Leu Ala
185 190 195
Met Ala Leu Ile Gly Ala Val Ser Ser Tyr Ile Ser Tyr Gln Gln
200 205 210
Lys Lys Phe Cys Phe Ser Ile Gln Gln Gly Leu Asn Ala Asp Tyr
215 220 225
Val Lys Gly Glu Asn Leu Glu Ala Val Val Cys Glu Glu Pro Gln
230 235 240
Val Lys Tyr Ser Thr Leu His Thr Gln Ser Ala Glu Pro Pro Pro
245 250 255
Pro Pro Glu Pro Ala Arg Ile
260

<210> 22
<211> 172
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2806595CD1

<400> 22
Met Gly Leu Leu Leu Leu Val Pro Leu Leu Leu Leu Pro Gly Ser
1 5 10 15
Tyr Gly Leu Pro Phe Tyr Asn Gly Phe Tyr Tyr Ser Asn Ser Ala
20 25 30
Asn Asp Gln Asn Leu Gly Asn Gly His Gly Lys Asp Leu Leu Asn
35 40 45
Gly Val Lys Leu Val Val Glu Thr Pro Glu Glu Thr Leu Phe Thr
50 55 60
Tyr Gln Gly Ala Ser Val Ile Leu Pro Cys Arg Tyr Arg Tyr Glu
65 70 75
Pro Ala Leu Val Ser Pro Arg Arg Val Arg Val Lys Trp Trp Lys
80 85 90
Leu Ser Glu Asn Gly Ala Pro Glu Lys Asp Val Leu Val Ala Ile
95 100 105
Gly Leu Arg His Arg Ser Phe Gly Asp Tyr Gln Gly Arg Val His
110 115 120
Leu Arg Gln Asp Lys Glu His Asp Val Ser Leu Glu Ile Gln Asp
125 130 135
Leu Arg Leu Glu Asp Tyr Gly Arg Tyr Arg Cys Glu Val Ile Asp
140 145 150

Gly Leu Glu Asp Glu Ser Gly Leu Val Glu Leu Glu Leu Arg Gly
155 160 165
Glu Met Leu Thr Gly Thr Gly
170

<210> 23
<211> 571
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2850987CD1

<400> 23
Met Thr Arg Ala Gly Asp His Asn Arg Gln Arg Gly Cys Cys Gly
1 5 10 15
Ser Leu Ala Asp Tyr Leu Thr Ser Ala Lys Phe Leu Leu Tyr Leu
20 25 30
Gly His Ser Leu Ser Thr Trp Gly Asp Arg Met Trp His Phe Ala
35 40 45
Val Ser Val Phe Leu Val Glu Leu Tyr Gly Asn Ser Leu Leu Leu
50 55 60
Thr Ala Val Tyr Gly Leu Val Val Ala Gly Ser Val Leu Val Leu
65 70 75
Gly Ala Ile Ile Gly Asp Trp Val Asp Lys Asn Ala Arg Leu Lys
80 85 90
Val Ala Gln Thr Ser Leu Val Val Gln Asn Val Ser Val Ile Leu
95 100 105
Cys Gly Ile Ile Leu Met Met Val Phe Leu His Lys His Glu Leu
110 115 120
Leu Thr Met Tyr His Gly Trp Val Leu Thr Ser Cys Tyr Ile Leu
125 130 135
Ile Ile Thr Ile Ala Asn Ile Ala Asn Leu Ala Ser Thr Ala Thr
140 145 150
Ala Ile Thr Ile Gln Arg Asp Trp Ile Val Val Val Ala Gly Glu
155 160 165
Asp Arg Ser Lys Leu Ala Asn Met Asn Ala Thr Ile Arg Arg Ile
170 175 180
Asp Gln Leu Thr Asn Ile Leu Ala Pro Met Ala Val Gly Gln Ile
185 190 195
Met Thr Phe Gly Ser Pro Val Ile Gly Cys Gly Phe Ile Ser Gly
200 205 210
Trp Asn Leu Val Ser Met Cys Val Glu Tyr Val Leu Leu Trp Lys
215 220 225
Val Tyr Gln Lys Thr Pro Ala Leu Ala Val Lys Ala Gly Leu Lys
230 235 240
Glu Glu Glu Thr Glu Leu Lys Gln Leu Asn Leu His Lys Asp Thr
245 250 255
Glu Pro Lys Pro Leu Glu Gly Thr His Leu Met Gly Val Lys Asp
260 265 270
Ser Asn Ile His Glu Leu Glu His Glu Gln Glu Pro Thr Cys Ala
275 280 285
Ser Gln Met Ala Glu Pro Phe Arg Thr Phe Arg Asp Gly Trp Val
290 295 300
Ser Tyr Tyr Asn Gln Pro Val Phe Leu Ala Gly Met Gly Leu Ala
305 310 315
Phe Leu Tyr Met Thr Val Leu Gly Phe Asp Cys Ile Thr Thr Gly
320 325 330
Tyr Ala Tyr Thr Gln Gly Leu Ser Gly Ser Ile Leu Ser Ile Leu
335 340 345
Met Gly Ala Ser Ala Ile Thr Gly Ile Met Gly Thr Val Ala Phe
350 355 360
Thr Trp Leu Arg Arg Lys Cys Gly Leu Val Arg Thr Gly Leu Ile
365 370 375
Ser Gly Leu Ala Gln Leu Ser Cys Leu Ile Leu Cys Val Ile Ser
380 385 390
Val Phe Met Pro Gly Ser Pro Leu Asp Leu Ser Val Ser Pro Phe

```
                    395                           400                           405
Glu Asp Ile Arg Ser Arg Phe Ile Gln Gly Glu Ser Ile Thr Pro
                    410                           415                           420
Thr Lys Ile Pro Glu Ile Thr Thr Glu Ile Tyr Met Ser Asn Gly
                    425                           430                           435
Ser Asn Ser Ala Asn Ile Val Pro Glu Thr Ser Pro Glu Ser Val
                    440                           445                           450
Pro Ile Ile Ser Val Ser Leu Leu Phe Ala Gly Val Ile Ala Ala
                    455                           460                           465
Arg Ile Gly Leu Trp Ser Phe Asp Leu Thr Val Thr Gln Leu Leu
                    470                           475                           480
Gln Glu Asn Val Ile Glu Ser Glu Arg Gly Ile Ile Asn Gly Val
                    485                           490                           495
Gln Asn Ser Met Asn Tyr Leu Leu Asp Leu Leu His Phe Ile Met
                    500                           505                           510
Val Ile Leu Ala Pro Asn Pro Glu Ala Phe Gly Leu Leu Val Leu
                    515                           520                           525
Ile Ser Val Ser Phe Val Ala Met Gly His Ile Met Tyr Phe Arg
                    530                           535                           540
Phe Ala Gln Asn Thr Leu Gly Asn Lys Leu Phe Ala Cys Gly Pro
                    545                           550                           555
Asp Ala Lys Glu Val Arg Lys Glu Asn Gln Ala Asn Thr Ser Val
                    560                           565                           570
Val
```

<210> 24
<211> 455
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3557211CD1

<400> 24

```
Met Asp Pro Thr Gly Asn Ser Ala Thr Pro Gln Ile Leu Glu Leu
 1                  5                   10                  15
Lys Trp Ser His Ile Glu Trp Ser Gln Thr Glu Tyr Ile Cys Glu
                    20                  25                  30
Asn Val Gly Leu Leu Pro Leu Glu Ile Ile Arg Arg Gly Tyr Ser
                    35                  40                  45
Met Asp Ser Ala Phe Val Gly Ile Lys Val Asn Gln Val Ser Ala
                    50                  55                  60
Ala Val Gly Lys Asp Phe Thr Val Ile Pro Ser Lys Leu Ile Gln
                    65                  70                  75
Phe Asp Pro Gly Met Ser Thr Lys Met Trp Asn Ile Ala Ile Thr
                    80                  85                  90
Tyr Asp Gly Leu Glu Glu Asp Asp Glu Val Phe Glu Val Ile Leu
                    95                  100                 105
Asn Ser Pro Val Asn Ala Val Leu Gly Thr Lys Thr Lys Ala Ala
                    110                 115                 120
Val Lys Ile Leu Asp Ser Lys Gly Gly Gln Cys His Pro Ser Tyr
                    125                 130                 135
Ser Ser Asn Gln Ser Lys His Ser Thr Trp Glu Lys Gly Ile Trp
                    140                 145                 150
His Leu Leu Pro Pro Gly Ser Ser Ser Ser Thr Thr Ser Gly Ser
                    155                 160                 165
Phe His Leu Glu Arg Arg Pro Leu Pro Ser Ser Met Gln Leu Ala
                    170                 175                 180
Val Ile Arg Gly Asp Thr Leu Arg Gly Phe Asp Ser Thr Asp Leu
                    185                 190                 195
Ser Gln Arg Lys Leu Arg Thr Arg Gly Asn Gly Lys Thr Val Arg
                    200                 205                 210
Pro Ser Ser Val Tyr Arg Asn Gly Thr Asp Ile Ile Tyr Asn Tyr
                    215                 220                 225
His Gly Ile Val Ser Leu Lys Leu Glu Asp Asp Ser Phe Pro Thr
                    230                 235                 240
His Lys Arg Lys Ala Lys Val Ser Ile Ile Ser Gln Pro Gln Lys
```

245 250 255
Thr Ile Lys Val Ala Glu Leu Pro Gln Ala Asp Lys Val Glu Ser
260 265 270
Thr Thr Asp Ser His Phe Pro Arg Gln Asp Gln Leu Pro Ser Phe
275 280 285
Pro Lys Asn Cys Thr Leu Glu Leu Lys Gly Leu Phe His Phe Glu
290 295 300
Glu Gly Ile Gln Lys Leu Tyr Gln Cys Asn Gly Ile Ala Trp Lys
305 310 315
Ala Trp Ser Pro Gln Thr Lys Asp Val Glu Asp Lys Ser Cys Pro
320 325 330
Ala Gly Trp His Gln His Ser Gly Tyr Cys His Ile Leu Ile Thr
335 340 345
Glu Gln Lys Gly Thr Trp Asn Ala Ala Ala Gln Ala Cys Arg Glu
350 355 360
Gln Tyr Leu Gly Asn Leu Val Thr Val Phe Ser Arg Gln His Met
365 370 375
Arg Trp Leu Trp Asp Ile Gly Gly Arg Lys Ser Phe Trp Ile Gly
380 385 390
Leu Asn Asp Gln Val His Ala Gly His Trp Glu Trp Ile Gly Gly
395 400 405
Glu Pro Val Ala Phe Thr Asn Gly Arg Arg Gly Pro Ser Pro Arg
410 415 420
Ser Lys Leu Gly Lys Ser Cys Val Leu Val Gln Arg Gln Gly Lys
425 430 435
Trp Gln Thr Lys Asp Cys Arg Arg Ala Lys Pro His Asn Tyr Val
440 445 450
Cys Ser Arg Lys Leu
455

<210> 25
<211> 437
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4675668CD1

<400> 25
Met Pro Lys Phe Lys Ala Ala Arg Gly Val Gly Gly Gln Glu Lys
1 5 10 15
His Ala Pro Leu Ala Asp Gln Ile Leu Ala Gly Asn Ala Val Arg
20 25 30
Ala Gly Val Arg Glu Lys Arg Arg Gly Arg Gly Thr Gly Glu Ala
35 40 45
Glu Glu Glu Tyr Val Gly Pro Arg Leu Ser Arg Arg Ile Leu Gln
50 55 60
Gln Ala Arg Gln Gln Gln Glu Glu Leu Glu Ala Glu His Gly Thr
65 70 75
Gly Asp Lys Pro Ala Ala Pro Arg Glu Arg Thr Thr Arg Leu Gly
80 85 90
Pro Arg Met Pro Gln Asp Gly Ser Asp Asp Glu Asp Glu Glu Trp
95 100 105
Pro Thr Leu Glu Lys Ala Ala Thr Met Thr Ala Ala Gly His His
110 115 120
Ala Glu Val Val Val Asp Pro Glu Asp Glu Arg Ala Ile Glu Met
125 130 135
Phe Met Asn Lys Asn Pro Pro Ala Arg Arg Thr Leu Ala Asp Ile
140 145 150
Ile Met Glu Lys Leu Thr Glu Lys Gln Thr Glu Val Glu Thr Val
155 160 165
Met Ser Glu Val Ser Gly Phe Pro Met Pro Gln Leu Asp Pro Arg
170 175 180
Val Leu Glu Val Tyr Arg Gly Val Arg Glu Val Leu Ser Lys Tyr
185 190 195
Arg Ser Gly Lys Leu Pro Lys Ala Phe Lys Ile Ile Pro Ala Leu
200 205 210

```
Ser Asn Trp Glu Gln Ile Leu Tyr Val Thr Glu Pro Glu Ala Trp
                215                 220                 225
Thr Ala Ala Ala Met Tyr Gln Ala Thr Arg Ile Phe Ala Ser Asn
                230                 235                 240
Leu Lys Glu Arg Met Ala Gln Arg Phe Tyr Asn Leu Val Leu Leu
                245                 250                 255
Pro Arg Val Arg Asp Asp Val Ala Glu Tyr Lys Arg Leu Asn Phe
                260                 265                 270
His Leu Tyr Met Ala Leu Lys Lys Ala Leu Phe Lys Pro Gly Ala
                275                 280                 285
Trp Phe Lys Gly Ile Leu Ile Pro Leu Cys Glu Ser Gly Thr Cys
                290                 295                 300
Thr Leu Arg Glu Ala Ile Ile Val Gly Ser Ile Ile Thr Lys Cys
                305                 310                 315
Ser Ile Pro Val Leu His Ser Ser Ala Ala Met Leu Lys Ile Ala
                320                 325                 330
Glu Met Glu Tyr Ser Gly Ala Asn Ser Ile Phe Leu Arg Leu Leu
                335                 340                 345
Leu Asp Lys Lys Tyr Ala Leu Pro Tyr Arg Val Leu Asp Ala Leu
                350                 355                 360
Val Phe His Phe Leu Gly Phe Arg Thr Glu Lys Arg Glu Leu Pro
                365                 370                 375
Val Leu Trp His Gln Cys Leu Leu Thr Leu Val Gln Arg Tyr Lys
                380                 385                 390
Ala Asp Leu Ala Thr Asp Gln Lys Glu Ala Leu Leu Glu Leu Leu
                395                 400                 405
Arg Leu Gln Pro His Pro Gln Leu Ser Pro Glu Ile Arg Arg Glu
                410                 415                 420
Leu Gln Ser Ala Val Pro Arg Asp Val Glu Asp Val Pro Ile Thr
                425                 430                 435
Val Glu
```

<210> 26
<211> 2893
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 398269CB1

<400> 26

```
agtggctgag tcgggggcgg gctgggaggg ctgtcggtgg gccagtctgc gtacgacggc   60
ccgtcccctg cgcacggacg ccgggaagaa gggggtgggg ccacgtttgc gtccgcgcca  120
tcaggcccga gatagcggcg aggtccgctt tcagtgtatg gttttccctg ccaaacggtt  180
ctgcttggtg ccatccatgg agggcgtgcg ctgggccttt tcctgcggca cttggctgcc  240
gagccgagcc gaatggctgc tggcagtgcg atcgattcag cccgaggaga aggagcgcat  300
tggccagttc gtctttgccc gggacgctaa ggcagccatg gctggtcgtc tgatgataag  360
gaaattagtt gcagagaaat tgaatatccc ttggaatcat attcgtttgc aaagaactgc  420
aaaaggaaaa ccagttcttg caaaggactc atcgaatcct tacccgaatt tcaactttaa  480
catctctcat caaggagact atgcagtgct tgctgctgaa cctgagctgc aagttggaat  540
tgatataatg aagactagtt ttccaggtcg tggttcaatt ccagaattct ttcatattat  600
gaaaagaaag tttaccaaca aagaatggga aacaatcaga agctttaagg atgagtggac  660
tcagctggat atgttttata ggaattgggc acttaaggaa agcttcataa aagccattgg  720
tgttggacta ggatttgaat tgcagcggct tgaatttgat ctatctccat taaacttgga  780
tataggccaa gtttataaag aaacacgttt attcctggat ggagaggaag aaaaagaatg  840
ggcatttgag gaaagcaaaa tagatgagca ccattttgtt gcagttgctc ttaggaaacc  900
cgatggatct agacatcagg atgttccatc tcaggatgat tccaaaccaa cccagaggca  960
atttactatt ctcaacttta atgatttaat gtcatctgcc gttcccatga cacctgaaga 1020
tccttcattt tgggactgtt tttgcttcac agaagaaatt ccaatacgaa atggtacaaa 1080
gtcatgatga ttccctgagt aacaaaggga aatgaaaact gtttgtgatc ttccgtattc 1140
actgaaaaat aaatgcttgt ttagtatcaa attttatttc acgaaagttt ttttaaagaa 1200
cagaaacttt tccaattaaa aaaaaaaagc agacttctgg ttcaagatag ctcactggaa 1260
tacatgttta cctctttctt tcctaaattg cattgaattg ataggaagga tggcggaatc 1320
ttaaagtgat acatgctaac tgtagaaaaa aatagaaaat gcacataagc aaaaggaaac 1380
atttaaatgc tatctttcaa agataactac tcttaaaacc ttgagtatct tttcagacct 1440
ttttcttggc aaatgaatcc atattgacat atttgatttt tttaaaaaca tggaaacgta 1500
```

ctgttttgta aattcttttt aactgcacat ctactgttca taaatatacc tctgtaacat 1560
aactttttgt ggttctaatg tactctgttg tatagctaat acaagttagg atgcttttgg 1620
ccagaggtaa cagtgtccaa atataattgg cctaagtaac ctaggaaatt gtttgacata 1680
acacagggtt caggggtgtc attaaagaca cacttttttt gccttgacct cagttggttt 1740
gttttgcctt aggctattcc acctctcgat cacaagagct gctgctataa cttccagttt 1800
tacatctttg taaaattgta tctcaaaggc agaaaaggcc atttcgtctc tcatttgttt 1860
tatccatgag gaagattttt aacaaaagcc tccagaagat ttcccctcag tttccattga 1920
cttagatcag gttacagaga aaggcaatgt ctgacatttt tggtctctgt tagaagtaga 1980
ctctgttgaa aagaaagaag ctaagctagg tgtgaagaat ggaattggaa gcccactgcc 2040
ttcccataag aaaggtttac cataatttac tcactttttt ctgtgttaga cattttgatt 2100
atctgcagtt tattactaca agcagtggca gagtgaatgt ccttgtacat tttgagttac 2160
atgcttaatt atgtccttga gaaagtttct aaaagtggaa tgattggttt gactggttca 2220
tagggcttta attatacaat ttacccctct aattagtact atatgtatgt gacttccctc 2280
cccctgccag aatactcctt ggtcaattgt aggtattctt tttggtttaa tttttgccaa 2340
tgtaattaaa aaatggtatg tcatttttaa aatttgtatt tctttcatta caaataagat 2400
tgttatgtca gtattgttat tggcttttcg tattcctctt aacgtgaacc gtctgttcat 2460
tgtttttacc tgttttcgtt ttagcaagta gtacttaatt taaagtgtga acttaatata 2520
taagatgcca ggaccatcat attgatgaca aaaatctatt atgtggttgt agtgcatgtc 2580
ttggagttaa actcttaagt atcaagaagt aaaataattt cattgtttct gttactaaca 2640
ttaaaagtgt gcaaattgta taaaattagg tacttcgatt agtaaaagaa aaaaatttga 2700
atgttttttt ttttatttta ttaagcatgc ccagttatgc caagcatagt aaataaaggt 2760
caagtagcat ttataataga ggaagtattg ttatccctag catgagtgta atggtgatat 2820
gaaaaacttt gtcttgtcat tataataata aaaaaatgaa catttattat ggaatttcaa 2880
aaaaaaaaaa aaa 2893

<210> 27
<211> 2276
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1258888CB1

<400> 27
gcgagtggag cggaggaccc gagcggctga ggagagagga ggcggcggct tagctgctac 60
ggggtccggc cggcgccctc ccgagggggg ctcaggagga ggaaggagga cccgtgcgag 120
aatgcctctg ccctggagcc ttgcgctccc gctgctgctc tcctgggtgg caggtggttt 180
cgggaacgcg gccagtgcaa ggcatcacgg gttgttagca tcggcacgtc agcctggggt 240
ctgtcactat ggaactaaac tggcctgctg ctacggctgg agaagaaaca gcaagggagt 300
ctgtgaagct acatgcgaac ctggatgtaa gtttggtgag tgcgtgggac caaacaaatg 360
cagatgcttt ccaggataca ccgggaaaac ctgcagtcaa gatgtgaatg agtgtggaat 420
gaaaccccgg ccatgccaac acagatgtgt gaatacacac ggaagctaca agtgcttttg 480
cctcagtggc cacatgctca tgccagatgc tacgtgtgtg aactctagga catgtgccat 540
gataaactgt cagtacagct gtgaagacac agaagaaggg ccacagtgcc tgtgtccatc 600
ctcaggactc cgcctggccc caaatggaag agactgtcta gatattgatg aatgtgcctc 660
tggtaaagtc atctgtccct acaatcgaag atgtgtgaac acatttggaa gctactactg 720
caaatgtcac attggtttcg aactgcaata tatcagtgga cgatatgact gtatagatat 780
aaatgaatgt actatggata gccatacgtg cagccaccat gccaattgct tcaataccca 840
agggtccttc aagtgtaaat gcaagcaggg atataaaggc aatggacttc ggtgttctgc 900
tatccctgaa aattctgtga aggaagtcct cagagcacct ggtaccatca aagacagaat 960
caagaagttg cttgctcaca aaaacagcat gaaaaagaag gcaaaaatta aaaatgttac 1020
cccagaaccc accaggactc ctacccctaa ggtgaacttg cagcccttca actatgaaga 1080
gatagtttcc agaggcggga actctcatgg aggtaaaaaa gggaatgaag agaaaatgaa 1140
agaggggctt gaggatgaga aaagagaaga gaaagccctg aagaatgaca tagaggagcg 1200
aagcctgcga ggagatgtgt tttccctaa ggtgaatgaa gcaggtgaat tcggcctgat 1260
tctggtccaa aggaaagcgc taacttccaa actggaacat aaagcagatt taaatatctc 1320
ggttgactgc agcttcaatc atgggatctg tgactggaaa caggatagag aagatgattt 1380
tgactggaat cctgctgatc gagataatgc tattggcttc tatatggcag ttccggcctt 1440
ggcaggtcac aagaaagaca ttggccgatt gaaacttctc ctacctgacc tgcaacccca 1500
aagcaacttc tgtttgctct ttgattaccg gctggccgga gacaaagtcg ggaaacttcg 1560
agtgtttgtg aaaaacagta acaatgccct ggcatgggag aagaccacga gtgaggatga 1620
aaagtggaag acagggaaaa ttcagttgta tcaaggaact gatgctacca aaagcatcat 1680
ttttgaagca gaacgtggca agggcaaaac cggcgaaatc gcagtggatg gcgtcttgct 1740
tgtttcaggc ttatgtccag atagcctttt atctgtggat gactgaatgt tactatcttt 1800
atatttgact ttgtatgtca gttccctggt ttttttgata ttgcatcata ggacctctgg 1860
catttagaa ttactagctg aaaaattgta atgtaccaac agaaatatta ttgtaagatg 1920

cctttcttgt ataagatatg ccaatatttg ctttaaatat catatcactg tatcttctca 1980
gtcatttctg aatctttcca cattatatta taaaatatgg aaatgtcagt ttatctcccc 2040
tcctcagtat atctgatttg tataagtaag ttgatgagct tctctctaca acatttctag 2100
aaaatagaaa aaaaagcaca gagaaatgtt taactgtttg actcttatga tacttcttgg 2160
aaactatgac atcaaagata gacttttgcc taagtggctt agctgggtct ttcatagcca 2220
aacttgtata tttaaattct ttgtaataat aatatccaaa tcatcaaaaa aaaaaa 2276

<210> 28
<211> 2016
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1375891CB1

<400> 28
gaaaaggtac ccgcgagaga cagccagcag ttctgtggag cagcggtggc cggctaggat 60
gggctgtctc tggggtctgg ctctgcccct tttcttcttc tgctgggagg ttggggtctc 120
tgggagctct gcaggcccca gcacccgcag agcagacact gcgatgacaa cggacgacac 180
agaagtgccc gctatgactc tagcaccggg ccacgccgct ctggaaactc aaacgctgag 240
cgctgagacc tcttctaggg cctcaacccc agccggcccc attccagaag cagagaccag 300
gggagccaag agaatttccc ctgcaagaga gaccaggagt ttcacaaaaa catctcccaa 360
cttcatggtg ctgatcgcca cctccgtgga gacatcagcc gccagtggca gccccgaggg 420
agctggaatg accacagttc agaccatcac aggcagtgat cccgaggaag ccatctttga 480
caccctttgc accgatgaca gctctgaaga ggcaaagaca ctcacaatgg acatattgac 540
attggctcac acctccacag aagctaaggg cctgtcctca gagagcagtg cctcttccga 600
cggcccccat ccagtcatca ccccgtcacg ggcctcagag agcagcgcct cttccgacgg 660
cccccatcca gtcatcaccc cgtcacgggc ctcagagagc agcgcctctt ccgacggccc 720
ccatccagtc atcaccccgt catggtcccc gggatctgat gtcactctcc tcgctgaagc 780
cctggtgact gtcacaaaca tcgaggttat taattgcagc atcacagaaa tagaaacaac 840
aacttccagc atccctgggg cctcagacat agatctcatc cccacggaag ggtgaaggc 900
ctcgtccacc tccgatccac cagctctgcc tgactccact gaagcaaaac cacacatcac 960
tgaggtcaca gcctctgccg agaccctgtc cacagccggc accacagagt cagctgcacc 1020
tcatgccacg gttgggaccc cactccccac taacagcgcc acagaaagag aagtgacagc 1080
acccggggcc acgaccctca gtggagctct ggtcacagtt agcaggaatc ccctggaaga 1140
aacctcagcc ctctctgttg agacaccaag ttacgtcaaa gtctcaggag cagctccggt 1200
ctccatagag gctgggtcag cagtgggcaa aacaacttcc tttgctggga gctctgcttc 1260
ctcctacagc ccctcggaag ccgccctcaa gaacttcacc ccttcagaga caccgaccat 1320
ggacatcgca accaaggggc ccttccccac cagcagggac cctcttcctt ctgtccctcc 1380
gactacaacc aacagcagcc gagggacgaa cagcacctta gccaagatca caacctcagc 1440
gaagaccacg atgaagcccc aacagccacg cccacgactg cccggacgag gccgaccaca 1500
gacgtgagtg caggtgaaaa tggaggtttc ctcctcctgc ggctgagtgt ggcttccccg 1560
gaagacctca ctgaccccag agtggcagaa aggctgatgc agcagctcca ccgggaactc 1620
cacgcccacg cgcctcactt ccaggtctcc ttactgcgtg tcaggagagg ctaacggaca 1680
tcagctgcag ccaggcatgt cccgtatgcc aaaagagggt gctgccccta gcctgggccc 1740
ccaccgacag actgcagctg cgttactgtg ctgagaggta cccagaaggt tcccatgaag 1800
ggcagcatgt ccaagcccct gaccccagat gtggcaacag gaccctcgct cacatccacc 1860
ggagtgtatg tgtggggagg ggcttcacct gttcccagag gtgtccttgg actcaccttg 1920
gcacatgttc tgtgtttcag taaagagaga cctgatcacc catctgtgtg cttccatcct 1980
gcattaaaat tcactcagtg tggcccaaaa aaaaaa 2016

<210> 29
<211> 2520
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1524355CB1

<400> 29
caagatggcg gctggcggag ctgtcgctgc ggcgcccgag tgccggcttc tcccctacgc 60
gctacacaag tggagctcct tttcctccac ctaccttccc gagaacattt tagtggacaa 120
accaaatgac caatcttcaa gatggtcttc agagagcaac tatcctcccc agtacttgat 180
tctaaagctc gaaaggcctg ctatagttca gaatatcaca tttggaaaat atgagaaaac 240
tcatgtttgc aatttgaaga aatttaaagt ctttggtgga atgaatgaag aaaatatgac 300

agagctgttg tccagtggct taaagaatga ttataacaaa gaaacattca ccttgaagca 360
taaaattgat gaacagatgt tcccttgtcg attcattaaa atagttccac tcttgtcctg 420
gggacccagc tttaacttta gcatctggta tgttgaactt agtggcattg atgatcctga 480
tatagtacaa cctgtgtctca actggtatag caagtaccgt gaacaggaag ctattcgcct 540
ttgcctaaaa cacttcagac aacacaacta tacagaagct tttgagtcac tgcaaaagaa 600
aaccaagatt gcactggaac atcccatgtt aacagatatt catgacaagc tggtgttgaa 660
gggtgatttt gatgcttgcg aagagttgat tgaaaaggct gtaaatgatg gcttgttcaa 720
tcagtatatc agtcaacagg aatataagcc acgatggagt caaatcattc ccaaaagtac 780
caaaggtgat ggggaagata accgtccagg aatgagagga ggccatcaga tggttattga 840
tgttcaaaca gagactgttt atttgtttgg tggctgggat ggaacacaag atcttgctga 900
cttctgggcg tacagtgtga aggagaacca gtggacatgt atctctagag acactgaaaa 960
agagaatggt cctagtgcca gatcgtgtca taaaatgtgc attgatattc aacggaggca 1020
aatctacaca ttggggcgtt acttggattc ctctgtgagg aacagcaaat ctctgaaaag 1080
tgacttctat cgttatgaca ttgatacaaa cacatggatg ttactaagtg aggatactgc 1140
tgctgatgga gggccgaaat tggtgtttga tcatcagatg tgtatggact cagaaaaaca 1200
tatgatctac acttttggtg gtagaatttt gacttgtaat ggcagcgtag atgacagcag 1260
agccagtgaa ccacaattca gtggcttgtt tgctttcaac tgtcaatgtc aaacctggaa 1320
acttcttcga gaggactcct gtaatgctgg gcctgaggac atccagtctc gaataggaca 1380
ctgcatgtta ttccactcaa aaaatcgttg cttatatgta tttggtggcc agcgatcaaa 1440
gacctatttg aatgatttct ttagttatga tgtggactct gatcatgtag acataatatc 1500
agatggcacc aagaaagact ctgggatggt tccaatgaca ggatttacac agagagcaac 1560
tattgatcca gaactgaatg aaatacacgt cttatctgga ctcagcaaag ataaggaaaa 1620
gagggaagaa aatgttagaa attcattctg gatttatgac attgtgagga atagttggtc 1680
ttgtgtctat aagaatgatc aagctgcaaa ggataatcca actaaaagtc ttcaggaaga 1740
agaaccatgt ccaaggtttg cccatcagct tgtatacgat gagctacaca aggttcatta 1800
cttatttggt gggaatccag gaaaatcttg ctctccaaag atgagattag atgacttctg 1860
gtcactgaag ttgtgtagac cttcaaaaga ttatttactg aggcattgca agtacctcat 1920
aagaaaacac aggtttgaag aaaaggccca agtggatccc cttagtgctc tgaaatattt 1980
acaaaatgat ctttatataa ctgtggatca ttcagaccca gaagagacaa aagagtttca 2040
gctcctggca tcagctctat tcaaatctgg ttcagatttt acagctctgg gcttttctga 2100
tgtggatcac acctatgctc aaagaactca gctctttgac accttagtaa atttctttcc 2160
tgacagcatg actcctccta aaggcaacct ggtagacctc atcacactgt aactgaagag 2220
tcactggaca cagaaatgga aaacaggagt cgattttccg tcttttggat tgcagctcca 2280
ctgactgaca gtaaagctgc agtgattgag gactgcacca gagttctgaa gggatcttaa 2340
ccatcacaag tttttaccct cttccttcat gcctgacctc aaccccgctc tcctcatcct 2400
attcctaaat taggctaata aagtgaaatt ggtatacttt ccagttaaat atatatatat 2460
atatattttt tcttacttta tcttttaaga attaatgagt ataaaagcaa aattaggcag 2520

<210> 30
<211> 1954
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1598937CB1

<400> 30
ccgagttatg cagctccggg cggcaagggg tcctgtcgag ggcgcggtc cataagggtt 60
gaggccagag gcgcgtacct ctgggcgccg agctctggag tggaggctgg ggttcggagt 120
gcgtcatctg gaagcaggca ccccggccgg caggcagagt cacggtggca gcattgagag 180
ttggacaccc gggtccttga agtgatctct aggccccagc cccaaatccg ccaccattcc 240
gtgctgcggg gacaccatgg ctccagaaga ggacgctgga ggggaggcct taggggcag 300
tttctgggag gctggcaact acaggcgcac ggtacagcgg gtggaggacg ggcaccggct 360
gtgcggggac ctggtcagct gcttccagga gcgcgcccgc atcgagaagg cttatgccca 420
gcagttggct gactgggccc gaaagtggag ggggaccgtg gagaaggcc cccagtatgg 480
cacactggag aaggcctggc atgccttttt cacggcggct gagcggctga gcgcgctgca 540
cctggaggtg cgggagaagc tgcaaggca ggacagtgag cgggtgcgcg cctggcagcg 600
gggggctttc caccggcctg tgctgggcgg cttccgcgag agccgggcgg ccgaggacgg 660
cttccgcaag gcccagaagc cctggctgaa gaggctgaag gaggttgagg cttccaagaa 720
aagctaccac gcagcccgga aggatgagaa gaccgcccag acgagggaga gccacgcaaa 780
ggcagacagc gccgtctccc aggagcagct gcgcaaactg caggaacggg tggaacgctg 840
tgccaaggag gccgagaaga caaaagctca gtatgagcag acgctggcag agctgcatcg 900
ctacactcca cgctacatgg aggacatgga acaggccttt gagacctgcc aggccgccga 960
gcgccagcgg cttcttttct tcaaggatat gctgctcacc ttacaccagc acctggacct 1020
ttccagcagt gagaagttcc atgaactcca ccgtgacttg caccagggca ttgaggcagc 1080
cagtgacgaa gaggatctgc gctggtggcg cagcacccac gggccaggca tggccatgaa 1140

ctggccacag ttcgaggagt ggtccttgga cacacagagg acaatcagcc ggaaagagaa 1200
gggtggccgg agccctgatg aggttaccct gaccagcatt gtgcctacaa gagatggcac 1260
cgcacccccca ccccagtccc cggggtcccc aggcacgggg caggatgagg agtggtcaga 1320
tgaagagagt ccccggaagg ctgccaccgg ggttcgggtg agggcactct atgactacgc 1380
tggccaggaa gctgatgagc tgagcttccg agcaggggag gagctgctga agatgagtga 1440
ggaggacgag cagggctggt gccaaggcca gttgcagagt ggccgcattg gcctgtaccc 1500
tgccaactac gtggagtgtg tgggcgcctg agtgtcctga cagcccttct gcaacgttta 1560
cccaccctgg ttcagagccc agcttctcct ggagagccgg accctcaggg ccctgaaccg 1620
tcgctctctg gctgctcctc tgtcccttga gggaggaagt cctgggaccc agggagggga 1680
ggggcctttg tctagggaag ggactggtag ggaagggacg agtctaggct gagggcaaga 1740
tgggaggtca gaggtgacag aagcgttcag gggtgcctgg gcctccccag gagctgtgga 1800
ctcagttcct gacctctgct ttggggttcc tggggtgggc ttggggtgag tgtagttctg 1860
gcctagcagc accctcttgt ggcttgttct agcgtgtatt aaaacttgac acacacccac 1920
acacaaaaac aaaaacacca aaaaaaaaaa aaaa 1954

<210> 31
<211> 1817
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1725801CB1

<400> 31
gacgcggtga ggagacggcc cacggcgccc gcgggctggg gcggtcgctt cttccttctc 60
cgtggcctac gagggtcccc agcctgggta aagatggccc catggccccc gaagggccta 120
gtcccagctg tgctctgggg cctcagcctc ttcctcaacc tcccaggacc tatctggctc 180
cagccctctc cacctcccca gtcttctccc ccgcctcagc cccatccgtg tcatacctgc 240
cggggactgg ttgacagctt taacaagggc ctggagagaa ccatccggga caactttgga 300
ggtggaaaca ctgcctggga ggaagagaat ttgtccaaat acaaagacag tgagacccgc 360
ctggtagagg tgctggaggg tgtgtgcagc aagtcagact tcgagtgcca ccgcctgctg 420
gagctgagtg aggagctggt ggagagctgg tggtttcaca agcagcagga ggcccccggac 480
ctcttccagt ggctgtgctc agattccctg aagctctgct gccccgcagg caccttcggg 540
ccctcctgcc ttccctgtcc tggggggaaca gagaggccct gcggtggcta cgggcagtgt 600
gaaggagaag ggacacgagg gggcagcggg cactgtgact gccaagccgg ctacgggggt 660
gaggcctgtg gccagtgtgg ccttggctac tttgaggcag aacgcaacgc cagccatctg 720
gtatgttcgg cttgttttgg cccctgtgcc cgatgctcag gacctgagga atcaaactgt 780
ttgcaatgca agaagggctg ggccctgcat cacctcaagt gtgtagacat tgatgagtgt 840
ggcacagagg gagccaactg tggagctgac caattctgcg tgaacactga gggctcctat 900
gagtgccgag actgtgccaa ggcctgccta ggctgcatgg gggcagggcc aggtcgctgt 960
aagaagtgta gccctggcta tcagcaggtg ggctccaagt gtctcgatgt ggatgagtgt 1020
gagacagagg tgtgtccggg agagaacaag cagtgtgaaa acaccgaggg cggttatcgc 1080
tgcatctgtg ccgagggcta caagcagatg gaaggcatct gtgtgaagga gcagatccca 1140
gagtcagcag gcttcttctc agagatgaca gaagacgagt tggtggtgct gcagcagatg 1200
ttctttggca tcatcatctg tgcactggcc acgctggctg ctaagggcga cttggtgttc 1260
accgccatct tcattggggc tgtggcggcc atgactggct actggttgtc agagcgcagt 1320
gaccgtgtgc tggagggctt catcaagggc agataatcgc ggccaccacc tgtaggacct 1380
cctcccaccc acgctgcccc cagagcttgg gctgccctcc tgctggacac tcaggacagc 1440
ttggtttatt tttgagagtg gggtaagcac ccctacctgc cttacagagc agcccaggta 1500
cccaggcccg ggcagacaag gcccctgggg taaaaagtag ccctgaaggt ggataccatg 1560
agctcttcac ctggcgggga ctggcaggct tcacaatgtg tgaatttcaa aagtttttcc 1620
ttaatggtgg ctgctagagc tttggcccct gcttaggatt aggtggtcct cacaggggtg 1680
gggccatcac agctccctcc tgccagctgc atgctgccag ttcctgttct gtgttcacca 1740
catccccaca ccccattgcc acttatttat tcatctcagg aaataaagaa aggtcttgga 1800
aagttaaaaa aaaaaaa 1817

<210> 32
<211> 2694
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1730482CB1

<400> 32

```
gacctagtgt gagcataatg gaaaaaacac aatcacttcc tacacgacca ccaacttttc   60
ctccaaccat tccaccagca aaagaagtat gtaaggcggc caaggctgac ctggtattta  120
tggtggatgg atcctggagc attggagatg aaaatttcaa taagatcatc agctttctat  180
acagcactgt tggagccctg aacaagattg gcacagatgg aacccaagtt gcaatggttc  240
agttcactga tgatcccaga acagaattta aactaaatgc ttacaaaacc aaagagactc  300
ttcttgatgc aattaaacac atttcataca aaggaggaaa tacaaaaaca ggaaaagcaa  360
ttaagtatgt tcgagatacc ttgttcactg cagagtcagg tacaagaagg ggcatcccaa  420
aggttatcgt ggttataact gatggaagat cacaagatga tgtgaacaaa atctccaggg  480
agatgcaatt agatggctat agcatttttg caattggtgt ggccgatgca gattactcgg  540
agttggttag cattggcagt aagcccagcg cacgccatgt cttctttgtg gatgactttg  600
acgcctttaa gaaaatcgaa gatgagttaa ttacttttgt ctgcgaaaca gcatcagcaa  660
cctgtccagt ggtacacaag gatggcattg atcttgcagg atttaagatg atggaaatgt  720
ttggtttggt tgaaaaagat ttttcatcag tggaaggggt ttctatggag cctggtacct  780
tcaatgtgtt tccatgttac caactccata aagatgccct ggtttcccag ccaaccaggt  840
acttgcaccc agaaggattg ccctccgact acacaatcag ttttctattc cggattcttc  900
ctgacactcc acaggagcca tttgctcttt gggagatttt aaataaaaat tctgacccat  960
tggttggggt tattttagac aatggtggga aaactctaac atatttcaac tatgaccaga 1020
gtggggattt tcaaactgtt actttcgaag gacctgaaat taggaaaatt ttttatggaa 1080
gctttcacaa gctacacatt gttgtcagtg agactttggt caaagtggtt attgactgca 1140
agcaagtggg tgagaaggca atgaacgcat cagctaatat cacgtcagat ggtgtagaag 1200
tgctagggaa aatggttcga tcaagaggac caggtggaaa ctctgcaccg ttccagttac 1260
agatgtttga tattgtttgc tccacatcat gggccaatac agacaaatgc tgtgaacttc 1320
caggcctgag agatgatgag tcttgcccag accttcccca ttcctgctcc tgttctgaaa 1380
ccaatgaagt ggctctggga ccagcgggcc caccaggtgg tccaggactc cgaggaccaa 1440
agggccagca aggtgaaccg ggtccaaagg gaccagatgg ccctcggggt gaaattggtc 1500
tgccaggacc tcagggtcca cctggacctc aaggaccaag tggtctgtcc attcaaggaa 1560
tgcccggaat gccaggagaa aaaggagaga aaggagatac tggccttcca ggtccacagg 1620
gtatcccagg aggcgttggt tcaccaggac gtgatggctc accaggccag aggggccttc 1680
cgggaaagga tggatcctcg ggacctccag gaccaccagg gccaataggc attcctggca 1740
cccctggagt cccagggatc acaggaagca tgggaccgca aggcgccctg ggaccacctg 1800
gtgtccctgg agcaaagggg gaacgaggag agcggggtga cctgcagtct caagccatgg 1860
tgagatcagt ggcgcgtcaa gtatgcgaac agctcatcca gagtcacatg gccaggtaca 1920
ctgccatcct caaccagatt cccagccact cctcatccat ccggactgtc caagggcctc 1980
ctggggagcc tgggaggcca ggctcacctg gagcccctgg tgaacaagga cccccaggca 2040
caccaggctt ccccggaaat gcaggcgtgc cagggacccc aggagaacga ggtctaactg 2100
gtatcaaagg agaaaaagga aatccaggcg ttggaaccca aggtccaaga ggcccccctg 2160
gaccagcagg accttcaggg gagagtcggc ctggcagccc tgggccccct ggctctcctg 2220
gaccaagagg cccaccaggt catctggggg ttcctggacc ccaaggtcct tctggccagc 2280
ctggatattg tgacccctca tcatgttctg cctatggtgt gagagctccc catccagatc 2340
agccagagtt caccccctgtc caagatgagc tggaagccat ggaactgtgg ggccctggag 2400
tctgatagcc tcaggagaaa tttgaagacc aactgcaaga actcttaagg aatcttgttt 2460
gagaaaatgt tgttatgtgg tttgtatgct actttttgggg ggcagggctc atttcagcag 2520
cctaaatctc ctccttggat aatgttaata ttattattat tattaacaaa aaatatatat 2580
ttttaaaaag ttcccttaat ctatgacatg gtagcaatga tttccctttg gtgtcttaat 2640
ggcatgtcag ataatttgtt tttccagaga agagagctca aagaggaatt ggga       2694
```

<210> 33
<211> 1149
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1810058CB1

<400> 33

```
cagtatctgg gtccagcctg cagccttagg gtccaggtga tgttaccgtg tgtgtggccc   60
ttcttcacag tggcctccta gaaaaacaag accctgactc aaagaacacc tctcactaca  120
ttcagagtct gtcatctgaa ccatgaggat ctggtggctt ctgcttgcca ttgaaatctg  180
cacagggaac ataaactcac aggacacctg caggcaaggg caccctggaa tccctgggaa  240
ccccggtcac aatggtctgc ctggaagaga tggacgagac ggagcgaagg gtgacaaagg  300
cgatgcagga gaaccaggac gtcctggcag cccggggaag gatgggacga gtggagagaa  360
gggagaacga ggagcagatg gaaaagttga agcaaaaggc atcaaaggtg atcaaggctc  420
aagaggatcc ccaggaaaac atggccccaa ggggcttgca gggcccatgg gagagaaagg  480
cctccgagga gagactgggc ctcaggggca gaaggggaat aagggtgacg tgggtcccac  540
tggtcctgag ggccaaggg gcaacattgg gcctttgggc ccaactggtt taccgggccc  600
catgggccct attggaaagc ctggtcccaa gggagaagct ggacccacgg ggccccaggg  660
```

```
tgagccagga gtccggggaa taagaggctg gaaaggagat cgaggagaga aagggaaaat 720
cggtgagact ctagtcttgc caaaaagtgc tttcactgtg ggctcacgg tgctgagcaa 780
gtttccttct tcagatgtgc ccattaaatt tgataagatc cacatcactg ttttctccag 840
gaatgttcag gtgtctttgg tcaaaaacgg agtaaaaata ctgcacacca gagatgctta 900
cgtgagctct gaggaccagg cctctggcag cattgtcctg cagctgaagc tcggggatga 960
gatgtggctg caggtgacag gaggagagag gttcaatggc ttgtttgctg atgaggacga 1020
tgacacaact ttcacagggt tccttctgtt cagcagccag tgacagagga gagtttataa 1080
atctgccaga ccatccatca gaatcagctt gggatgaact tattcagatg gttttacttt 1140
attaattca                                                         1149
```

<210> 34
<211> 1215
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2040679CB1

<400> 34

```
gaagaactag catgtatgta ttatctccag tggaatttat aattctacaa cttttattta   60
ttcaggccat ttccagcagt ttaaaaggtt tcctttcagc tatgagactg gctcatagag  120
gctgtaatgt tgatacacca gtttcaacgc tcacaccagt gaagacttca gaatttgaaa  180
actttaaaac taaaatggtt atcacatcca aaaaagacta tcctctaagt aagaattttc  240
catattcctt ggaacatctt cagacttctt actgtgggct tgtccgagtt gatatgcgta  300
tgctttgctt aaaaagcctt aggaaattag acttgagtca caaccatata aaaaagcttc  360
cagctacaat tggagacctc atacaccttc aagaacttaa cctgaatgac aatcacttgg  420
agtcatttag tgtagccttg tgtcattcta cactccagaa gtcacttcgg agtttggacc  480
tcagcaagaa caaaatcaag gcactccctg tgcagttttg ccagctccag gaacttaaga  540
atttaaaact tgacgataat gaattgattc aatttccttg caagatagga caactaataa  600
accttcgctt tttgtcagca gctcgaaata agcttccatt tttgcctagt gaatttagaa  660
atttatccct tgaatacttg gatctttttg gaaatacttt tgaacaacca aaagtccttc  720
cagtaataaa gctgcaagca ccattaactt tattggaatc ttctgcacga accatattac  780
ataataggat tccatatggc tctcatatca ttccattcca tctctgccaa gatttggata  840
ccgcaaaaat ttgtgtttgt ggaagattct gtctgaactc tttcattcaa ggaactacta  900
ccatgaatct gcattctgtt gcccacactg tggtcttagt agataatttg ggtggtactg  960
aagcacctat tatctcttat ttctgttctc taggctgtta tgttaattcc tctgatatgt 1020
taaagtaatg ggtgagacca gaaaaagaaa tttcaataac agatcagttt ggggtgcatg 1080
tatgattttg cagcgtcaaa ttggagtaag ggaagatttc tgtatacttg ctggagagga 1140
ggaatgtgta tagttactca tttagatgac tccaaaactt ttattaaaac caattttagt 1200
tttaaaaaaa aaaaa                                                  1215
```

<210> 35
<211> 1300
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2960051CB1

<400> 35

```
ttctgcacaa agaactgaaa ggcatttatc cccaagggag gcagttattt tagattttac   60
taagaagttc agcaaatact tttcaacatt cccttctgtc ctttctttgt ttttaaagaa  120
agctctgatt ttgtttcatt ttcagctgga gacttaaatg acaccaagca aagcctactt  180
agtttagatc tccagaaatt ggctggtgga aaaaaatcaa acatgaagat tgcagttttg  240
ttttgttttt ttctgcttat catttttcaa actgactttg gaaaaaatga agaaattcct  300
aggaagcaaa ggaggaagat ctaccacaga aggttgagga aaagttcaac ctcacacaag  360
cacagatcaa acagacagct tggaattccg caaacaacag tttttacacc agtagcaaga  420
cttcctattg ttaactttga ttatagcatg gaggaaaagt ttgaatcctt ttcaagtttt  480
cctggagtag aatcaagcta taatgtgtta ccaggaaaga agggacactg tttggtaaag  540
ggcataacca tgtacaacaa agctgtgtgg tcgcctgagc cctgcactac ctgcctctgc  600
tcagatggaa gagttctttg tgatgaaacc atgtgccatc cccagaggtg cccccaaaca  660
gttatacctg aaggggaatg ctgccctgtc tgctccgcta ctggtacaga gatttagcta  720
agcaaaatat cagtgtgtga ttaatcttta acttccattt gtttttgtta ctaattttag  780
attaaaatta tgatacatta gtcagatctg agtacttaaa atattggcaa aatgctgatt  840
aacatagaaa atatctggga aaatgtatgg taggggatat aaataataga ctgtggcttt  900
```

atagttctag ctctatcaga ttcagtaaac ttggatgaga ttacattcca catttgactc 960
tcagctttag agatatggta acagaatttc tacaacagat cctgaattct tattgcatta 1020
agggctctgc tttggtctat atgtgcatta tcccacttaa tccagtgcaa cgtgccttta 1080
tcaccttgaa gccagggtaa acaaaggaag agtgatttgc atctaaagag aacaaagccc 1140
caaccctctg gctataccca accactcaaa ggcagcacag gaacccacat cactgcttgg 1200
ataatcccag gaaaatgcag aaaaagtgta gcctgaagca tgattttctc atgtggcact 1260
tctgtgtgca ggagatcaca gcgcggtttt gttgctgcca 1300

<210> 36
<211> 1562
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3117318CB1

<400> 36
aaggccgggc gcggtaagag cgtctcgggg agtagggcaa ggcggccggg cccctcccat 60
tccgcctttt cttcagcgtc ctgcccgcgg cactggctgc gggtgccggg ccacctgcga 120
gtgtgcgcag ggactctgga cacccgcggc ggcgagctga gggagcagtc tccacgagga 180
cccaggcgga ccctctggcg ccatgcgcgc cctcccggc ctgctggagg ccagggcgcg 240
tacgccccgg ctgctcctcc tccagtgcct tctcgctgcc gcgcgcccaa gctcggcgga 300
cggcagtgcc ccagattcgg cttttacaag tccacctctc agagaagaaa taatggcaaa 360
taacttttcc ttggagagtc ataacatatc actgactgaa cattctagta tgccagtaga 420
aaaaaatatc actttagaaa ggccttctaa tgtaaatctc acatgccagt tcacaacatc 480
tggggatttg aatgcagtaa atgtgacttg gaaaaaagat ggtgaacaac ttgagaataa 540
ttatcttgtc agtgcaacag gaagcacctt gtatacccaa tacaggttca ccatcattaa 600
tagcaaacaa atgggaagtt attcttgttt ctttcgagag gaaaaggaac aaaggggaac 660
atttaatttc aaagtccctg aacttcatgg gaaaaacaag ccattgatct cttacgtagg 720
ggattctact gtcttgacat gtaaatgtca aaattgtttt cctttaaatt ggacctggta 780
cagtagtaat gggagtgtaa aggttcctgt tggtgttcaa atgaataaat atgtgatcaa 840
tggaacatat gctaacgaaa caaagctgaa gataacacaa cttttggagg aagatgggga 900
atcttactgg tgccgtgcac tattccaatt aggcgagagt gaagaacaca ttgagcttgt 960
ggtgctgagc tatttggtgc ccctcaaacc atttcttgta atagtggctg aggtgattct 1020
tttagtggcc accattctgc tttgtgaaaa gtacacacaa aagaaaaaga agcactcaga 1080
tgaggggaaa gaatttgagc agattgaaca gctgaaatca gatgatagca atggtataga 1140
aaataatgtc cccaggcata gaaaaaatga gtctctgggc cagtgaatac aaaacatcat 1200
gtcgagaatc attggaagat atacagagtt cgtatttcag ctttgtttat ccttcctgtt 1260
aagagcctct gagtttttag ttttaaaagg atgaaaagct tatgcaacat gctcagcagg 1320
agcttcatca acgatatatg tcagatctaa aggtatattt tcattctgta attatgttac 1380
ataaaagcaa tgtaaatcag aataaatatg ttagaccaga atacaaatta attatattct 1440
ggtcttcaaa ggacacacag aacagatatc agcagaatca cttaatactt catagaacaa 1500
aaatcactca aaacctgttt ataaccaaag aattcatgaa aaagaaagcc tttggccatt 1560
tg 1562

<210> 37
<211> 2801
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 2793
<223> a, t, c, g, or other

<220>
<221> misc_feature
<223> Incyte ID No: 3486992CB1

<400> 37
gactttgctt gaatgtttac attttctgct cgctgtccta catatcacaa tatagtgttc 60
acgttttgtt aaaactttgg ggtgtcagga gttgagcttg ctcagcaagc cagcatggct 120
aggatgagct ttgttatagc agcttgccaa ttggtgctgg gcctactaat gacttcatta 180
accgagtctt ccatacagaa tagtgagtgt ccacaacttt gcgtatgtga aattcgtccc 240
tggtttaccc cacagtcaac ttacagagaa gccaccactg ttgattgcaa tgacctccgc 300
ttaacaagga ttcccagtaa cctctctagt gacacacaag tgcttctctt acagagcaat 360

aacatcgcga agactgtgga tgagctgcag cagcttttca acttgactga actagatttc 420
tcccaaaaca actttactaa cattaaggag gtcgggctgg caaacctaac ccagctcaca 480
acgctgcatt tggaggaaaa tcagattacc gagatgactg attactgtct acaagacctc 540
agcaaccttc aagaactcta catcaaccac aaccaaatta gcactatttc tgctcatgct 600
tttgcaggct taaaaaatct attaaggctc cacctgaact ccaacaaatt gaaagttatt 660
gatagtcgct ggtttgattc tacacccaac ctggaaattc tcatgatcgg agaaaaccct 720
gtgattggaa ttctggatat gaacttcaaa cccctcgcaa atttgagaag cttagttttg 780
gcaggaatgt atctcactga tattcctgga aatgctttgg tgggtctgga tagccttgag 840
agcctgtctt tttatgataa caaactggtt aaagtccctc aacttgccct gcaaaaagtt 900
ccaaatttga aattcttaga cctcaacaaa aaccccattc acaaaatcca agaaggggac 960
ttcaaaaata tgcttcggtt aaaagaactg ggaatcaaca atatgggcga gctcgtttct 1020
gtcgaccgct atgccctgga taacttgcct gaactcacaa agctggaagc caccaataac 1080
cctaaactct cttacatcca ccgcttggct ttccgaagtg tccctgctct ggaaagcttg 1140
atgctgaaca acaatgcctt gaatgccatt taccaaaaga cagtcgaatc cctccccaat 1200
ctgcgtgaga tcagtatcca tagcaatccc ctcaggtgtg actgtgtgat ccactggatt 1260
aactccaaca aaaccaacat ccgcttcatg gagcccctgt ccatgttctg tgccatgccg 1320
cccgaatata aagggcacca ggtgaaggaa gttttaatcc aggattcgag tgaacagtgc 1380
ctcccaatga tatctcacga cagcttccca aatcgtttaa acgtggatat cggcacgacg 1440
gttttcctag actgtcgagc catggctgag ccagaacctg aaatttactg ggtcactccc 1500
attggaaata agataactgt ggaaacccttt tcagataaat acaagctaag tagcgaaggt 1560
accttggaaa tatctaacat acaaattgaa gactcaggaa gatacacatg tgttgcccag 1620
aatgtccaag ggcagacac tcgggtggca acaattaagg ttaacgggac ccttctggat 1680
ggtacccagg tgctaaaaat atacgtcaag cagacagaat cccattccat cttagtgtcc 1740
tggaaagtta attccaatgt catgacgtca aacttaaaat ggtcgtctgc caccatgaag 1800
attgataacc ctcacataac atatactgcc agggtcccag tcgatgtcca tgaatacaac 1860
ctaacgcatc tgcagccttc cacagattat gaagtgtgtc tcacagtgtc caatattcat 1920
cagcagactc aaaagtcatg cgtaaatgtc acaaccaaaa atgccgcctt cgcagtggac 1980
atctctgatc aagaaaccag tacagccctt gctgcagtaa tggggtctat gtttgccgtc 2040
attagccttg cgtccattgc tgtgtacttt gccaaaagat ttaagagaaa aaactaccac 2100
cactcattaa aaaagtatat gcaaaaaacc tcttcaatcc cactaaatga gctgtaccca 2160
ccactcatta acctctggga aggtgacagc gagaaagaca aagatggttc tgcagacacc 2220
aagccaaccc aggtcgacac atccagaagc tattacatgt ggtaactcag aggatatttt 2280
gcttctggta gtaaggagca caaagacgtt tttgctttat tctgcaaaag tgaacaagtt 2340
gaagactttt gtatttttga ctttgctagt ttgtggcaga gtggagagga cgggtggata 2400
tttcaaattt ttttagtata gcgtatcgca agggtttgac acggctgcca gcgactctag 2460
gcttccagtc tgtgtttggt ttttattctt atcattatta tgattgttat tatattatta 2520
ttttatttta gttgttgtgc taaactcaat aatgctgttc taactacagt gctcaataaa 2580
atgattaatg acaggatggg gttcccctgt gcttttacca gtagcatgac cccttctgaa 2640
gccatccgta gaaagtactt tgtccccaaa aagcaacata cggtttgaac agcatgaaac 2700
tttgtagcat cgggctaaga ctttaactca gagcaaggca gactggtacc tcgttaagat 2760
gtagtgactg cggatgttta cactgaatga agntgcttaa t 2801

<210> 38
<211> 2597
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4568384CB1

<400> 38
ccagaggcaa agaggccagt gaggactgct ctgtgcagtt gtccaggcct gaagaaggtg 60
gcggtgattt gaatcagaag aggtggcatt ctcttcatta ggaggatatg gataatcaag 120
acctaacaaa ttgggctccg aaagatacct tcgtcatggg gaagagtgaa tggagagaag 180
ataatataaa atgccaaacg tatataaatt gaaaggggaa aattgagaga acttgcgatc 240
catggtcttg gtcttccata aaggggaact gggacatcca ctggagcaga gcacagattg 300
gcccaagagc cccaagactc ccactggcct ccgcagaggc cgacagtgta ttcgtcctgc 360
ggagattgtg gcttccctgt tagaaggaga ggagaacacc tgtggcaaac agaaaccaaa 420
ggaaaacaat ttaaagccaa aatttcaggc tttcaaggga gtaggctgtc tatatgaaaa 480
ggagtcaatg aaaaaatcct tgaaagacag tgttgcctct aacaataaag atcagaattc 540
catgaaacat gaggatccca gtatcatatc catggaagat gggtccccat atgttaatgg 600
ctcattaggt gaagtgactc catgccaaca tgcaaagaag gcgaatggcc caaactatat 660
tcagcctcaa aaaagacaga ccactttga aagccaggat cgcaaggcag tgtcccctag 720
cagttctgaa aagagaagta agaatcctat ttctaggcca ttagaaggta agaagtcctt 780
aagtcttagt gcaaagactc acaacatagg ctttgacaaa gacagctgcc atagtaccac 840
aaagacagaa gcttcacagg aagagcggtc tgattcaagc ggcctcacat ctctcaagaa 900

atcaccaaag gtctcatcca aggacactcg ggaaatcaaa actgatttct cactttctat 960
tagtaattcg tcagatgtga gtgctaaaga taagcatgct gaagacaatg agaagcgttt 1020
ggcagccttg gaagcgaggc aaaaagcaaa agaagtgcag aagaagctgg tgcataatgc 1080
tctggcaaat ttggatggtc atccagagga taagccaacg cacatcatct tcggttctga 1140
cagtgaatgt gaaacagagg agacatcgac tcaggagcag agccatccag gagaggaatg 1200
ggtgaaagag tctatgggta aaacatcagg gaagctgttt gatagcagtg atgatgacga 1260
atctgattct gaagatgaca gtaataggtt caaaattaaa cctcagtttg agggcagagc 1320
tggacagaag ctcatggatt tacagtcgca ctttggcacc gatgacagat tccgcatgga 1380
ctctcgattt ctagaaactg acagtgaaga ggaacaggaa gaggtaaatg aaaagaaaac 1440
tgctgaggaa gaagagcttg ctgaagaaaa aaagaaagcc ctgaatgttg tacaaagtgt 1500
tttgcaaatc aacttaagca attctacaaa cagaggatca gtagctgcta agaaatttaa 1560
ggacatcata cattatgatc caacgaagca agaccatgcc acttacgaaa gaaaaagaga 1620
tgataaacca aaagaaagta aagcaaaacg aaaaaagaaa agggaggaag ctgagaaact 1680
acctgaggtg tctaaagaaa tgtattataa tattgctatg gatctgaaag aaatattcca 1740
aactacaaaa tataccagtg aaaaggaaga gggcacaccc tggaatgagg actgtggtaa 1800
agagaaacct gaggaaatcc aggaccctgc agctctgacc agtgacgctg agcagcccag 1860
cgggttcacg ttctcttttt ttgattcaga cactaaagac ataaaggaag agacctacag 1920
agttgaaaca gtgaaacctg gaaagattgt ctggcaggaa gaccctcgtt tacaagacag 1980
cagttcagaa gaggaagatg ttactgaaga aacagatcac agaaactcca gtcctggaga 2040
agcatcatta cttgagaaag agaccactag atttttcttt ttctctaaga atgatgaacg 2100
acttcaaggt tctgacttat tctggagagg agtaggaagt aatatgagca ggaactcttg 2160
ggaggccaga acaaccaacc tgcgtatgga ttgtcgaaag aaacataaag acgcaaaaag 2220
gaaaatgaaa ccaaaataat aaatgtcagc tggttttgat actgaatgtg aacaaggctc 2280
acctaaggaa actgacccag aaaacagttt tagctgacaa agaagaaatt tcagagtgaa 2340
ggaattttaa aaatctggct gacggaatat cattctggtt gccatctttt tctgtggaac 2400
tcctctgcat ttcttcctaa gtaattactt caaaaattaa attcaacttc ttataaagga 2460
agaacaagat agtccttgaa aatacttttt gtatataatc tctttgccct ctatcctgag 2520
taactaatgg acatcttctc atgcaaggtt tatatgaagc ctttttaaat aaatgagtca 2580
aagcaaaaaa aaaaaaa 2597

<210> 39
<211> 2641
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4586187CB1

<400> 39
ctgggaagaa agctatcagc accaactcag aactctccac cttcagatca gacattctag 60
atctccgtca gcaacttcgt gagattacag aaaaaaccag caagaacaag gatacgctgg 120
agaagttaca ggcgagcggg gatgctctgg tggacaggca gagtcaattg aaagaaactt 180
tggagaataa ctctttcctc atcaccactg taaacaaaac cctccaggcg tataatggct 240
atgtcacgaa tctgcagcaa gataccagcg tgctccaggg caatctgcag aaccaaatgt 300
attctcataa tgtggtcatc atgaacctca acaacctgaa cctgacccag gtgcagcaga 360
ggaacctcat cacgaatctg cagcggtctg tggatgacac aagccaggct atccagcgaa 420
tcaagaacga ctttcaaaat ctgcagcagg tttttcttca agccaagaag gacacggatt 480
ggctgaagga gaaagtgcag agcttgcaga cgctggctgc caacaactct gcgttggcca 540
aagccaacaa cgacaccctg gaggatatga acagccagct caactcattc acaggtcaga 600
tggagaacat caccactatc tctcaagcca acgagcagaa cctgaaagac ctgcaggact 660
tacacaaaga tgcagagaat agaacagcca tcaagttcaa ccaactggag gaacgcttcc 720
agctctttga gacggatatt gtgaacatca ttagcaatat cagttacaca gcccaccacc 780
tgcggacgct gaccagcaat ctaaatgaag tcaggaccac ttgcacagat acccttacca 840
aacacacaga tgatctgacc tccttgaata ataccctggc caacatccgt ttggattctg 900
tttctctcag gatgcaacaa gatttgatga ggtcgaggtt agacactgaa gtagccaact 960
tatcagtgat tatggaagaa atgaagctag tagactccaa gcatggtcag ctcatcaaga 1020
attttacaat actacaaggt ccaccgggcc ccaggggtcc aagaggtgac agaggatccc 1080
aggaccccc tggcccaact ggcaacaagg gacagaaagg agagaagggg gagcctggac 1140
cacctggccc tgcgggtgag agaggcccaa ttggaccagc tggtcccccc ggagagcgtg 1200
gcggcaaagg atctaaaggc tcccagggcc ccaaaggctc ccgtggttcc cctgggaagc 1260
ccggccctca gggccccagt ggggacccag gcccccggg cccaccaggc aaagagggac 1320
tccccggccc tcagggccct cctggcttcc agggacttca gggcaccgtt ggggagcctg 1380
gggtgcctgg acctcgggga ctgccaggct tgcctggggt accaggcatg ccaggcccca 1440
agggcccccc cggccctcct ggcccatcag gagcggtggt gcccctggcc ctgcagaatg 1500
agccaacccc ggcaccggag gacaatagct gcccgcctca ctggaagaac ttcacagaca 1560
aatgctacta tttttcagtt gagaaagaaa tttttgagga tgcaaagctt ttctgtgaag 1620

```
acaagtcttc acatcttgtt ttcataaaca ctagagagga acagcaatgg ataaaaaaac 1680
agatggtagg gagagagagc cactggatcg gcctcacaga ctcagagcgt gaaaatgaat 1740
ggaagtggct ggatgggaca tctccagact acaaaaattg gaaagctgga cagccggata 1800
actggggtca tggccatggg ccaggagaag actgtgctgg gttgatttat gctgggcagt 1860
ggaacgattt ccaatgtgaa gacgtcaata acttcatttg cgaaaaagac agggagacag 1920
tactgtcatc tgcattataa cggactgtga tgggatcaca tgagcaaatt ttcagctctc 1980
aaaggcaaag gacactcctt tctaattgca tcaccttctc atcagattga aaaaaaaaaa 2040
gcactgaaaa ccaattactg aaaaaaaatt gacagctagt gtttttttacc atccgtcatt 2100
acccaaagac ttgggaacta aaatgttccc cagggtgata tgctgatttt cattgtgcac 2160
atggactgaa tcacatagat tctcctccgt cagtaaccgt gcgattatac aaattatgtc 2220
ttccaaagta tggaacactc caatcagaaa aaggttatca ttggtcgttg agttatggga 2280
agaacttaag catatactgt gtaaacagtg ccatacattt ctaaaatccc aagtgtagga 2340
aaaatatgca gacatacaga tatataggcc aactattagt aataatatga aatatactta 2400
aagagctttt aaaactttgt atttttgtac aaaatatttg tcttttacaa tttttttcct 2460
tttttttttt ttgtcatttt accgacataa tacatggagc caaagaaaac aataatggta 2520
ctaataaaaa ctcctagggt ttcctgtcag atttaattct acccagtggc aaagaatttt 2580
ttcaattgtg gctttaaaaa aataattaaa tatacatgta tatatatata aaaaaaaaaa 2640
a                                                                 2641
```

<210> 40
<211> 914
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 401801CB1

<400> 40

```
cagaggtgta ctctacatga tgcacaaatg tgatatgtct ttttggtgaa gagcacgtga  60
tgaaaacagt aaataactga gaatagcaca aagctactag ggactcagat gattcaaata 120
ttgaagacta tgagtaatat taatcagcaa cttagtttgt tatcttcagt tatatgggag 180
gagtacatta ttctttgtta taggactacc tcacccttaa attgtaagtt ctttattagc 240
ttattgcatt ttccattctt aaagcagtag tttagtgttc tcttactgta tgacaattaa 300
aagtatactt aattgactac ataatgtgat agttaaaaaa tatatttaaa gtagcttttt 360
gaaagctttg ctgtttcccc cccttttttgt atataaaagc attagttgtc attattcatg 420
tgttctcatt acttttttaca aatgagaaca atgccttatg catgttgtga acaattacta 480
aaattttatg taaatttaac ctttattttt aattaatatg tgttaagata taacattatt 540
ttatctatta atatatgtat tttaatttac aggaaaacat ctttatgtat ggaggcagaa 600
ttgaaacaaa tgatggcaat gtcacagatg aattatgggt ttttaacata catagtcagt 660
catggagtac aaaaactcct actgttcttg gacatggtca gcagtatgct gtggagggac 720
attcagcaca tattatggag ttggatagta gagatgttgt catgatcata atatttggat 780
attctgcaat atatggttat acaagcagca tacaggaata ccatatctgt gagttactta 840
aaaattgtaa tttctttatt gattgggaat gtttttctct ttaataaaat cttcatatga 900
atttaaaaaa aaaa                                                   914
```

<210> 41
<211> 1006
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1721842CB1

<400> 41

```
ggcaaagaga actacaaatc ccagcgtcac ccgcggcctt gaagccccgc ccctgacaaa  60
ctgaaggtcc cggtaagcat cgcgtcagta cttatggcgc ctgccgggtt gtggtgacga 120
aagcagttgc catggagttg ctctgagtaa ccctgaggca gtgggacgcc aagactggag 180
aggaagcgac tgcggggagt atttccattt taaccggaaa caatccctga acccacagga 240
atgaatgcct aatggtggag tttcagccat cagtgacagg ctgaacccag actcccaggg 300
cacctgcttg cacctttgaa tgatggcctg aactatgaac aaacgggact atatgaacac 360
ttcggtacag gagccccctc ttgactactc cttcagaagc atccacgtca ttcaagatct 420
ggtaaatgag gagccaagga caggactacg accactgaag cgttcaaagt cggggaaatc 480
actgacccag tccctgtggc tgaataacaa tgttctcaat gatctgagag acttcaacca 540
ggtggcttca cagctgttgg agcacccaga gaacctggcc tggatcgacc tgtcctttaa 600
tgacctgact tccattgacc ctgtcctaac aactttcttc aacctgagtg tcctctatct 660
```

tcacggcaac agcatccagc gcctggggga ggtgaataag ctggctgtcc ttcctcggct 720
ccgtagcctg acactccatg ggaaccccat ggaggaagag aaagggtata ggcaatatgt 780
gctgtgcacc ctgtcccgta tcaccacgtt cgacttcagt ggggtcacca aagcagaccg 840
caccacagct gaagtctgga aacgcatgaa catcaagccc aagaaggcct ggaccaagca 900
gaatacactt tgaggctccc acgaccctag tagtcctaaa ggcctaagca tagacagcat 960
ggtttgacaa taaataattt gagctgttga gcagaaaaaa aaaaaa 1006

<210> 42
<211> 2582
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1833221CB1

<400> 42
gttaaattta gtactgaatg cagccttttg ttgtttaaaa aatttttttg aactcacagt 60
ctacatcagc atcagcatct gcgtcaccat ttcaatctgc atggtatagt gaatctgaga 120
taactcaggg agcacgctca agatcgcaga accagcaacg ggatcatgat tcaaaaagac 180
ctaaactttc ctgtacaaac tgtactacct cagctgggag aaatgttgga aatggtttaa 240
acacattatc agattcatct tggaggcata gtcaagttcc tagatcttca tcaatggtac 300
ttggatcatt tggaacagac ttaatgagag agaggagaga tttggagaga agaacagatt 360
cctctattag taatcttatg gattatagtc accgaagtgg tgatttcaca acttcatcat 420
atgttcaaga cagagttcct tcatattcac aaggagcaag accaaaagaa aactcaatga 480
gcactttaca gttgaataca tcatccacaa accaccaatt gccttctgaa catcagacca 540
tactaagttc tagggactcc agaaattctt taagatcaaa tttttcttca agagaatcag 600
aatcttcccg aagcaatacg cagcctggat tttcttacag ttcaagtaga gatgaagccc 660
caatcataag caattcagaa agggttgttt catctcaaag accatttcaa gaatcttctg 720
acaatgaagg taggcggaca acgaggagat tgctgtcacg catagcttct agcatgtcat 780
ctactttttt ttcacgaaga tctagtcagg attccttgaa tacaagatca ttgaattctg 840
aaaattctta cgtttctcca agaatcttga cagcttcaca gtcccgtagt aatgtaccat 900
cagcttctga agttcccgat aatagggcat ctgaagcttc tcagggattt cgatttctta 960
ggcgaagatg ggtttgtca tctcttagcc acaatcatag ctctgagtca gattcagaaa 1020
attttaacca agaatctgaa ggtagaaata caggaccatg gttatcttcc tcacttagaa 1080
atagatgcac acctttgttc tctagaagga ggcgagaggg aagagatgaa tcttcaagga 1140
tacctacctc tgatacatca tctagatctc atatttttag aagagaatca aatgaagtgg 1200
ttcaccttga agcacagaat gatcctcttg gagctgctgc caacagacca caagcatctg 1260
cagcatcaag cagtgccaca acaggtggct ctacatcaga ttcggctcaa ggtggaagaa 1320
atacaggaat atcagggatt cttcctggtt ccttattccg gtttgcagtc cccccagcac 1380
ttgggagtaa tttgaccgac aatgtcatga tcacagtaga tattattcct tcaggttgga 1440
attcagctga tggtaaaagt gataaaacta aaagtgcgcc ttcaagagat ccagaaagat 1500
tgcagaaaat aaaagagagc ctccttttag aggactcaga agaagaagaa ggtgacttat 1560
gtagaatttg tcaaatggca gctgcatcat catctaattt gctgatagag ccatgcaagt 1620
gcacaggaag tttgcagtat gtccaccaag actgtatgaa aaagtggtta caggccaaaa 1680
ttaactctgg ttcttcatta gaagctgtaa ccacctgtga actatgtaaa gagaagttgg 1740
agcttaacct ggaggatttt gatattcatg aactacatag agctcatgca aatgaacaag 1800
ctgagtatga gtttatcagc tctggtctct acctagtggt gttattgcac ttgtgcgaac 1860
aaagcttttc tgatatgatg ggaaatacaa atgaaccaag cacacgtgtc cgatttatta 1920
accttgcaag aactcttcag gcacatatgg aagatctcga aacttcagag gatgattccg 1980
aagaagacgg agaccataac aggacatttg atattgccta acttcatata agacagatgg 2040
atgatctgtg aacataagtg tttattaaaa atggcaatta aatataaatt actttgtgg 2100
gggaatgcct aataaataca ttgactatat ataaaatgaa tatatacata cacatgtatg 2160
cctgtatata tatattcatt ctccagtgtt gctgaattaa aattctgctg gactttttaa 2220
catagcaaat ccgatgttta taaactggta atcaaaaagg tttttcttt taggtgagtg 2280
ggaaagtatt acccttgttt taaatatcta agcaatgcct atcaaccctt ttttgtgtta 2340
tgattactgt agtcatattt atgaaaaaag gtttgtgttt tactcttgct agtgagaaaa 2400
gtgggacaaa atatactttt gaaataaaat gctatatggc acctaattat tttttctttt 2460
aaaatgcctt aagttgcagt ctcattttga taatcatttg cttccagtgt ttaaaaatta 2520
aaaaaagaat ggggagaagg ttatgagaag agcattatta agtttccaaa tttaatttga 2580
at 2582

<210> 43
<211> 2849
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2041168CB1

<400> 43
acgtgaccaa gagctgacgt gtgcagaagt ccttcttgtc ctggtcgttg ttcccgtctg 60
agtaccagct ccccactgcc ctgagggcgg gccggcctgc ggcggaggga aaaaggaaga 120
ggagaaggaa attgtcccga atccctgcag tctttctgta ggttgcggca caacgccagg 180
caaaagaaga ggaaggaatt taatcctaat cggtggaggt cgatttgagg gtctgctgta 240
gcaggtggct ccgcttgaag cgagggagga agtttcctcc gatcagtaga gattggaaag 300
attgttggga gtggcacacc actagggaaa agaagaaggg gcgaactgct tgtcttgagg 360
aggtcaaccc ccagaatcag ctcttgtggc cttgaagtgg ctgaagacga tcaccctcca 420
caggcttgag cccagtccca cagccttcct cccccagcct gagtgactac tctattcctt 480
ggtccctgct attgtcgggg acgattgcat gggctacgcc aggaaagtag gctgggtgac 540
cgcagcctgg tgattggggc tggcgcctgc tattgcattt atagactgac taggggaaga 600
aaacagaaca aggaaaaaat ggctgagggt ggatctgggg atgtggatga tgctggggac 660
tgttctgggg ccaggtataa tgactggtct gatgatgatg atgacagcaa tgagagcaag 720
agtatagtat ggtacccacc ttgggctcgg attgggactg aagctggaac cagagctagg 780
gccagggcaa gggccagggc tacccgggca cgtcgggctg tccagaaacg ggcttccccc 840
aattcagatg ataccgtttt gtcccctcaa gagctacaaa aggttctttg cttggttgag 900
atgtctgaaa agccttatat tcttgaagca gctttaattg ctctgggtaa caatgctgct 960
tatgcattta acagagatat tattcgtgat ctgggtggtc tcccaattgt cgcaaagatt 1020
ctcaatactc gggatcccat agttaaggaa aaggctttaa ttgtcctgaa taacttgagt 1080
gtgaatgctg aaaatcagcg caggcttaaa gtatacatga atcaagtgtg tgatgacaca 1140
atcacttctc gcttgaactc atctgtgcag cttgctggac tgagattgct tacaaatatg 1200
actgttacta atgagtatca gcacatgctt gctaattcca tttctgactt ttttcgttta 1260
ttttcagcgg gaaatgaaga aaccaaactt caggttctga aactcctttt gaatttggct 1320
gaaaatccag ccatgactag ggaactgctc agggcccaag taccatcttc actgggctcc 1380
ctctttaata agaaggagaa caaagaagtt attcttaaac ttctggtcat atttgagaac 1440
ataaatgata atttcaaatg ggaagaaaat gaacctactc agaatcaatt cggtgaaggt 1500
tcactttttt tctttttaaa agaatttcaa gtgtgtgctg ataaggttct gggaatagaa 1560
agtcaccatg atttttggt gaaagtaaaa gttggaaaat tcatggccaa acttgctgaa 1620
catatgttcc caaagagcca ggaataacac cttgattttg taatttagaa gcaacacaca 1680
ttgtaaacta ttcattttct ccaccttgtt tatatggtaa aggaatcctt tcagctgcca 1740
gttttgaata atgaatatca tattgtatca tcaatgctga tatttaactg agttggtctt 1800
taggtttaag atggataaat gaatatcact acttgttctg aaaacatgtt tgttgctttt 1860
tatctcgctg cctagattga aatattttgc tatttcttct gcataagtga cagtgaacca 1920
attcatcatg agtaagctcc cttctgtcat tttcattgat ttaatttgtg tatcatcaat 1980
aaaattgtat gttaatgctg gaaagaaaaa aagaagaaag aaaaaaacca tccctgtcct 2040
tcagtttata atctagttgg agagataaga aacgtacaaa ccaaaagata acagaatatc 2100
tgaagcatgt actcattgtc agatgttccc tctgagagca cagaggaggc aaaagcttct 2160
gtgggatgtg ctagtcggct aaagcttcac agaggaggtg gcaattgaaa atgagtcctg 2220
aatggggtag ggtggttagg gaattccatg agacaagaca agggggcat ggtgtgagaa 2280
aggcatggaa gtaggaaccc tcttcctatg acaggagatc attctgctta gagtggagag 2340
tgtggagagt gggagtagat aattttggaa agctgggtga agccagttgt ggagaattgt 2400
ttgaatatta tcccattgaa tacccagagc cactaaatct ttttttacta gaaataatt 2460
ggggtccata tgaaagtctc tattactgag tagtgtcaat gagggtgtgg caaaatggag 2520
cctttcacat cctagtggtg gccatttggt aatacagata taagccttaa actatgtaaa 2580
cccttgtcct aaggaagtaa ttgaataatt gcccaaagat tgtatgtatg aggctgttca 2640
tcccagcact gtctaagcta gtaaaaattg gaaacaattt aagtatctag cacattggat 2700
tggttataaa gcaaggaatg ttcacacagt aggatattat aagtatgctg atggaaatct 2760
atattgccag gaaaagctat tcattatgcg ttgtgaagtc agaaagtaaa aaagggtaga 2820
tagaagtatt cgaagtatag ttccatttt 2849

<210> 44
<211> 670
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2365794CB1

<400> 44
gaggcaagaa ttcggcacga ggggcgccgc gggcatttct tccactgccc gtctgaggga 60
acgctaagta gtgtgtccgg cgccgtgttc cagctccgcg ttgttccgcg agaaagcgag 120
aggccgagcc cgggctggtg cgatggccgc ggtggtggcc aagcgggaag ggccgccgtt 180

catcagcgag gcggccgtgc ggggcaacgc cgccgtcctg gattattgcc ggacctcggt 240
gtcagcgctg tcgggggcca cggccggcat cctcggcctc accggcctct acggcttcat 300
cttctacctg ctcgcctccg tcctgctctc cctgctcctc attctcaagg cgggaaggag 360
gtggaacaaa tatttcaaat cacggagacc tctctttaca ggaggcctca tcgggggcct 420
cttcacctac gtcctgttct ggacgttcct ctacggcatg gtgcacgtct actgaaatgg 480
gggcccgggg gactttttta aaaaaccaga tcgggaggac tgtggccagc aattaacacc 540
atgtagactt ccttagttct taagtggttg aattcgctgc ttgttctgta acgttataaa 600
taatttatat ctgaagacgg agagcctgta atattcttca gattaaatga agcgtgagac 660
aaaaaaaaaa 670

<210> 45
<211> 2364
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2618452CB1

<400> 45
ctcaatgcca acaggcacca ttcctccacc gacaacgctg aaggccacag ggtccaccca 60
cacagcccca ccaatgatgc caaccaccag tgggaccagc caagcctcaa gctccttcaa 120
cacagccaaa acctctacat ccctacattc acacacttcc tccacacacc atcctgaagt 180
caccccaact tctatcacca acatcaccct caaccccacc agtataggaa cctggacacc 240
cgtggcccac accacctcgg ccaccagcag caggctaacc acaccccttca ccacacattc 300
cccacctaca gggagcagtc ccatctcttc cacaggtcct atgactgcaa catcctttca 360
gaccaccact tattatacac caccatcaca ccctcagacc acacttccca ctcacgttcc 420
acctttctcc acctccttgg tgactccaag tactcacaca gtcatcatca ctacccacac 480
acagatggcc acttctgcct ccatccactc aacgccaaca ggcaccgttc ctccaccaac 540
aacgctcaag gccacagggt ccacccacac agccccacca atgacagtga ccaccagtgg 600
gaccagccaa acccacagct cattcagcac agctacagcc tcttcttcct tcatatcctc 660
ctcgtcttgg tcgtcttggc tgcctcagaa ctctagctca aggccaccgt catcacctat 720
caccacacaa ctcccccact tgagttctgc aaccactcct gtttccacaa ctaatcagct 780
gtcctcctca ttttctccca gtccttctgc cccctctact gtttcttctt atgtgccctc 840
ctcccactcc tctccccaga cttcatcgcc ttctgttggc acatcttcct ctttcgtgtc 900
cgcccccgtg cactccacaa ccctgagctc ggggtcacac tcctcattgt ccactcatcc 960
cacgactgca tcagtgtctg catctcctct ttttccttct tctccagctg cctctactac 1020
cattagggcc actctccccc acactatctc ctctcctttc accctctctg ctctactccc 1080
catatccact gttaccgtgt ctcccacccc atccagccac ctagcctcca gcaccattgc 1140
atttccgtcc acgcccagga ccacggccag cacccacacc gcccctgcct tctcctctca 1200
gtccaccacc tcgcggtcca cttctctcac cacccgagtt cccacatcag gctttgtgtc 1260
actcacctcg gggtgacgg gtatccccac ctctccagtc accaacctta ccaccaggca 1320
ccctggtccc accttgtcgc ctaccacacg gttcctgacc agctccctca ctgcccatgg 1380
aagcacccct gcttctgccc cggtatcttc tctcgggaca cctacgccca cctcacccgg 1440
ggtctgcagt gtgcgggagc agcaggagga gatcacgttc aaggggtgca tggcgaacgt 1500
gacggtaacc cgctgtgagg gcgcctgcat ttccgctgcc agcttcaaca tcatcaccca 1560
gcaggtggat gcccgctgca gctgctgccg ccccctccac tcctatgagc agcagctgga 1620
gctgccctgc cccgatccca gcacgcctgg ccggcggctc gtactcaccc tgcaggtgtt 1680
cagccactgc gtgtgcagct ctgtggcctg tggagactag cagggtcgct gcctgctctc 1740
ctggggctga aggactgcag atgacagaca ggaaaacacc caccagcccc cttcccgctt 1800
gtgccagcag ctgctttcct ggtcaccagg cctggccccc aagtgccctg ggccgtggct 1860
ccctggggca ccggttggag aggggctgcc aagcaggggc tcagactacc acactcctgc 1920
agaccctgag ccagcagaga gggactgagg cggacagtgg tcacggacct cccaggcaca 1980
cagggcactc ccgaccaccc ctgcccactg tccaacacct cccagcccct gaacttggcc 2040
ccagccctgc tgggcccaga accctgcaga tgaagccaca gagcaggcgc tcgaccagac 2100
ccatcagggg cgaggagggc acggaaacct gtgccgagat gggggcaaga ggcccaggca 2160
gccaccagca cagagaagag gagatcccca gagtcaggga gggcagaggg tggcagcgag 2220
ggcagggcag ccgcccccgc tcccagccag gcagaaggcc cccaccagca ccacacccat 2280
ccccagcagc ctgtccttgg gagagggcgt cacccggtca gagactccaa ataaaccggt 2340
tcttgtcaag gcacaaaaaa aaaa 2364

<210> 46
<211> 3600
<212> DNA
<213> Homo sapiens

<220>

<221> misc_feature
<223> Incyte ID No: 2622288CB1

<400> 46
gcagaggcgg cggggctcct cctcccgctc ctcctcggcc tccccttcgg gcgctctcgc 60
gctaactgtg ctcctccggg gccctccgcc tgctcccagc catggtggcc tggcgctcgg 120
cgttccttgt ctgcctcgct ttctccttgg ccaccctggt ccagcgagga tctggggact 180
ttgatgattt taacctggag gatgcagtga aagaaacttc ctcagtaaag cagccatggg 240
accacaccac caccaccaca accaataggc caggaaccac cagagctccg gcaaaacctc 300
caggtagtgg attggacttg gctgatgctt tggatgatca agatgatggc cgcaggaaac 360
cgggtatagg aggaagagag agatggaacc atgtaaccac cacgaccaag aggccagtaa 420
ccaccagagc tccagcaaat actttaggaa atgatttga cttggctgat gccctggatg 480
atcgaaatga tcgagatgat ggccgcagga aaccaattgc tggaggagga ggtttttcag 540
acaaggatct tgaagacata gtagggggtg gagaatacaa acctgacaag ggtaaaggtg 600
atggccggta cggcagcaat gacgaccctg gatctggcat ggtggcagag cctggcacca 660
ttgccggggt ggccagcgcc ctggccatgg ccctcatcgg tgccgtctcc agctacatct 720
cctaccagca gaagaagttc tgcttcagca ttcagcaggg tctcaacgca gactacgtga 780
agggagagaa cctggaagcc gtggtatgtg aggaacccca agtgaaatac tccacgttgc 840
acacgcagtc tgcagagccg ccgccgccgc ccgaaccagc ccggatctga gggccctgtc 900
cagctgcagg catgcacaat ggtgccaccg cttgtcaccc ggctcccccc acccсttcat 960
ttggacccgc agctgctgtg ctgctctgtg ccatcggctc cttgttggtc tgagtttccc 1020
ggatgagctc tgggtgtttg tgagtttggt ttctctgccc tgccccaagc gtgctgagac 1080
ttggtgccga aattcaagag ccagctctga tagaaagcca gcaccagcct cgggagctgc 1140
tgagccacca actcccaaag ccagcctgcc tccagcttta ctgagcacag gatgcggggg 1200
ccaagatgat gctgaggcct gatgacattt atgcttaggg gacaagagtt tgaactcaag 1260
ggactgtgac ccctgcacac tggagtggct cattgtggca ggtttctgcc aatagacagc 1320
ccctgacagt ggcctcaagg agctgcaggt gggggctca gcctgcaccc acttggagcc 1380
cctgcaagga gcgaaccggt cagcaccaag taacaccaca cacacgcagc acccaggatg 1440
atggtttcac ttcagtcttc cccatcccag gttttatgtt gctgggcttc cggagagccg 1500
gtccaagcgg aggctttcag tgatttaagt acaaacatgc atctcgtgat agtcctgcct 1560
tgagagctta ggaatcttcc ggataagtat gaagcaattc gtaggcctgt ttcccatctg 1620
attccatagg gggctgggtg tggccttcgg gttgacatga gaaaggtctt tagcaatcat 1680
ttctgcaccg gagatgagtt ttatcctgtg ttggggagag gtgctcaccc tccaccctgt 1740
gtccctgttt tggtagcaag agtgaccgat gtcaagaacg agcatcaaag ccagaatcct 1800
gcttgtttgc ttaaaaatgt aattgggggc ggcgggggag gagaggggaa agagacattc 1860
gcttggttta gtgaaacgca ggtgactttg tagctctgtg gtcagcctac ttgtctgctc 1920
tgagggagag tgcgtgggga gccatgctca ccgtggcaaa cacaggaacc ccatgactcg 1980
cccctcacct ggcgtggagc tgcctggttt gggctggagc agagctggtt tcctggaatg 2040
ttcctttggc ccacatatgg ttctgtcccg gtgagctctg ttgtcagagg ctcacgggac 2100
agaaccacat gctagggtct agggcccctg tctactgata gtcagtttgc tgtgtcagaa 2160
agcacttctg aaagcagata tgagtcacca gacaggcagg atcttacaaa actcacgggc 2220
ctctttggtc tgcatgatgg ccccatgcgt ttcataggct gtccactgag cgggattgtc 2280
tgctgagtgg gatgagccaa ctccagtttc ttaaggaaac cactggaatc tgcagccccc 2340
acatgcatct gtctaacgca tgcctcgtgt tcgttttgca aacatgcctg tggtggaggg 2400
tggtcagttg tagccctgtg cgtctcaagg ctgccttgtg aggccattcc cagtgcgtgc 2460
ccttgagctc cttaccaccc cttttcctgc tcggcccttt aatccctgac agacctggac 2520
tgtgtggctg aaggggggacc tgcagcactg cagaaatgcc tctgcgtggt gccatgaagg 2580
aaagaaacct tggcctggtc tcgagaagct tcccatgctt caggaagtta gtaagggtgg 2640
ggtggcttgc aggattggcc tgtttccagg gcctcccaca ctcattggcc agattgtgaa 2700
ctttgtcagg cttgtccctc cctgatacca agtatgtcga gaaccgatgg ccccaccctc 2760
tggctggtgc tgggccggag gtggctatgg aggattttgg catgcgtggc ctgtcgccac 2820
ctggacagcg tgacctcagg ggttgtccac tttaccttta tggtgaggcc tgtcggatgg 2880
ctaagtcctt gaaaccctag agctgtgacg tagaatatgt gctgtctgtg agaccgtgtt 2940
cccaggagca ctgactgcag ttgagagaga cccattttgc tctcccttac cgccccccgc 3000
cccgggtgct ttctgcacaa agcctagagc ctggcactca agcccaccgg tggcagctcc 3060
tagtgactgg acatgcctgg aagacccctc agccttctgt ttgcagaacg ttcatttcag 3120
gagcttctcc ttcccacaga catcttacac ttgctcgaca ctgccacctg cagaagcctg 3180
gcgggctctg gtcaccatgt gtctatctga aggttgcact ggccagcatg ggcctgtccc 3240
aagcgagagg ggagacacag tggactgaaa ggactggttg aaagtggcca atctctatca 3300
gcttaatttg gcagagaaaa tttgtaacaa ctctgagcac atgctgggtg aagtcacagc 3360
tcaaggaaag ataaagctgg gcggaaggag gtgtgcgtgg cttctggggt gggacccaga 3420
ggggaggctc tgggacaggg gctggggttc agtgccaggg ccctgaggaa gaaatgggga 3480
ctgatctcaa aattccagaa ttccctgtac atctgttcac gtgcttgtgt ccaggtgtga 3540
cttgtaaact gtctagtgtt tgcattaaat aaaatggcac cgagcagaaa aaaaaaaaaa 3600

<210> 47
<211> 1236

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2806595CB1

<400> 47
```
ttaatttccc cgaaatcaga ctgctgcctt ggaccgggac agctcgcggc ccccgagagc   60
tctagccgtc gaggagctgc ctggggacgt ttgccctggg gccccagcct ggcccgggtc  120
accctggcat gaggagatgg gcctgttgct cctggtcccg ttgctcctgc tgcccggctc  180
ctacggactg cccttctaca acggcttcta ctactccaac agcgccaacg accagaacct  240
aggcaacggt catggcaaag acctccttaa tggagtgaag ctggtggtgg agacacccga  300
ggagaccctg ttcacctacc aaggggccag tgtgatcctg ccctgccgct accgctacga  360
gccggccctg gtctccccgc ggcgtgtgcg tgtcaaatgg tggaagctgt cggagaacgg  420
ggccccagag aaggacgtgc tggtggccat cggggctgagg caccgctcct ttggggacta  480
ccaaggccgc gtgcacctgc ggcaggacaa agagcatgac gtctcgctgg agatccagga  540
tctgcggctg gaggactatg ggcgttaccg ctgtgaggtc attgacgggc tggaggatga  600
aagcggtctg gtggagctgg agctgcgggg tgagatgcta acggggactg ggtgacactg  660
ggacctgaga gcagagggga gaggaccaga gaaaacatcc agacctctgt gctttagaca  720
tttaaaagta cttaattctc aaaacaaccc acatggaagc tactgttgtg acccccattt  780
tacaggtgag aaaactgagg cacagagagg tcaagtaact tacctaaggt cacacagctt  840
gtaaacgaca gagctggggt ctgaacccaa gcacccagcc tctagaatct gttcccctct  900
acccgctgta attcacatct cattcagaga gaggaaaacc agagctggtc ccacagctta  960
ttagagacag agctgagatt taagcaaggt tagcgggtaa cacaagcgaa tgaggcagcc 1020
cactgtgaga cgatgggagt ggtggggact gggcacactc ctgagccctt gtcctgtgcc 1080
cagggagctc ccacactact caggcccagc tgatgatgcc acgccagaat gcagacccca 1140
ctgccaaatt ttctggcttg acaagataag cttatatttt tatgtgagat ctccagattt 1200
ttatgtaaaa acctcctttt taaaaacaaa acaaaa                           1236
```

<210> 48
<211> 3081
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2850987CB1

<400> 48
```
gcccggccca cggcggcggc ggcggcggcg gagagagctg gctcagggcg tccgctaggc   60
tcggacgacc tgctgagcct cccaaaccgc ttccataagg ctttgccttt ccaacttcag  120
ctacagtgtt agctaagttt ggaaagaagg aaaaaagaaa atccctgggc cccttttctt  180
ttgttctttg ccaaagtcgt cgttgtagtc tttttgccca aggctgttgt gtttttagag  240
gtgctatctc cagttccttg cactcctgtt aacaagcacc tcagcgagag cagcagcagc  300
gatagcagcc gcagaagagc cagcggggtc gcctagtgtc atgaccaggg cgggagatca  360
caaccgccag agaggatgct gtggatcctt ggccgactac ctgacctctg caaaattcct  420
tctctacctt ggtcattctc tctctacttg gggagatcgg atgtggcact ttgcggtgtc  480
tgtgtttctg gtagagctct atggaaacag cctccttttg acagcagtct acgggctggt  540
ggtggcaggg tctgttctgg tcctgggagc catcatcggt gactgggtgg acaagaatgc  600
tagacttaaa gtggcccaga cctcgctggt ggtacagaat gtttcagtca tcctgtgtgg  660
aatcatcctg atgatggttt tcttacataa acatgagctt ctgaccatgt accatggatg  720
ggttctcact tcctgctata tcctgatcat cactattgca aatattgcaa atttggccag  780
tactgctact gcaatcacaa tccaaaggga ttggattgtt gttgttgcag gagaagacag  840
aagcaaacta gcaaatatga atgccacaat acgaaggatt gaccagttaa ccaacatctt  900
agcccccatg gctgttggcc agattatgac atttggctcc ccagtcatcg gctgtggctt  960
tatttcggga tggaacttgg tatccatgtg cgtggagtac gtcctgctct ggaaggttta 1020
ccagaaaacc ccagctctag ctgtgaaagc tggtcttaaa gaagaggaaa ctgaattgaa 1080
acagctgaat ttacacaaag atactgagcc aaaacccctg gagggaactc atctaatggg 1140
tgtgaaagac tctaacatcc atgagcttga acatgagcaa gagcctactt gtgcctccca 1200
gatggctgag ccccttccgta ccttccgaga tggatgggtc tcctactaca accagcctgt 1260
gtttctggct ggcatgggtc ttgctttcct ttatatgact gtcctgggct ttgactgcat 1320
caccacaggg tacgcctaca ctcagggact gagtggttcc atcctcagta ttttgatggg 1380
agcatcagct ataactggaa taatgggaac tgtagctttt acttggctac gtcgaaaatg 1440
tggtttggtt cggacaggtc tgatctcagg attggcacag ctttcctgtt tgatcttgtg 1500
tgtgatctct gtattcatgc ctggaagccc cctggacttg tccgtttctc cttttgaaga 1560
tatccgatca aggttcattc aaggagagtc aattacacct accaagatac ctgaaattac 1620
```

aactgaaata tacatgtcta atgggtctaa ttctgctaat attgtcccgg agacaagtcc 1680
tgaatctgtg cccataatct ctgtcagtct gctgtttgca ggcgtcattg ctgctagaat 1740
cggtctttgg tcctttgatt taactgtgac acagttgctg caagaaaatg taattgaatc 1800
tgaaagaggc attataaatg gtgtacagaa ctccatgaac tatcttcttg atcttctgca 1860
tttcatcatg gtcatcctgg ctccaaatcc tgaagctttt ggcttgctcg tattgatttc 1920
agtctccttt gtggcaatgg gccacattat gtatttccga tttgcccaaa atactctggg 1980
aaacaagctc tttgcttgcg gtcctgatgc aaaagaagtt aggaaggaaa atcaagcaaa 2040
tacatctgtt gtttgagaca gtttaactgt tgctatcctg ttactagatt atatagagca 2100
catgtgctta ttttgtactg cagaattcca ataaatggct gggtgttttg ctctgttttt 2160
accacagctg tgccttgaga actaaaagct gtttaggaaa cctaagtcag cagaaattaa 2220
ctgattaatt tcccttatgt tgaggcatgg aaaaaaaatt ggaaaagaaa aactcagttt 2280
aaatacggag actataatga taacactgaa ttcccctatt tctcatgagt agatacaatc 2340
ttacgtaaaa gagtggttag tcacgtgaat tcagttatca tttgacagat tcttatctgt 2400
actagaattc agatatgtca gttttctgca aaactcactc ttgttcaaga ctagctaatt 2460
tatttttttg catcttagtt atttttaaaa acaaattctt caagtatgaa gactaaattt 2520
tgataactaa tattatcctt attgatccta ttgatcttaa ggtatttaca tgtatgtgga 2580
aaaacaaaac acttaactag aattctctaa taaggtttat ggtttagctt aaagagcacc 2640
tttgtatttt tattatcaga tggggcaaca tattgtatga agcatatgta gcacttcaca 2700
gcatggttat catgtaagct gcaggtagaa gcaaagctgt aaagtagatt tatcacacaa 2760
tgactgcata cagacttcaa atatgtcaat agtttggtca tagaacctag aagccaaaag 2820
ccacacagaa gggcaagaat cccaatttaa ctcatgttat catcattagt gatctgtgtt 2880
gtagaacatg agggtgtaag ccttcagcct ggcaagttac atgtagaaag cccacacttg 2940
tgaaggtttt gttttacaaa tcacttgatt taacacactc aggtagaata tttttatttt 3000
tactgtttta tacccagaag ttatttctac attgttctac agcaagaata ttcataaagg 3060
tggaccttgc aagtgcgtat a 3081

<210> 49
<211> 1825
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3557211CB1

<400> 49
cgtgttgaag gcatcagacc ctgacactga ggacgatcag ataatcttta aaattctaca 60
aggcccaaaa catggacatc tggagaacac aacaacaggt gaatttatcc atgagaaatt 120
tagccaaaag gacttaaaca gtaagactat tctttacatc ataaacccat ctttggaagt 180
aaattcagat accgtggaat ttcaaatcat ggaccccaca gggaactcgg ccactcctca 240
aattttggaa ctgaagtggt ctcatattga atggtcacag accgaatata tctgtgagaa 300
tgtgggtttg ttgcccttgg aaattatcag aaggggatat tccatggact cggcctttgt 360
gggtataaag gtcaaccaag tgtcagctgc agttggaaaa gatttcaccg tgattccatc 420
taaactgatt cagtttgacc caggaatgtc aactaagatg tggaatatag caattaccta 480
tgacggatta gaggaagatg atgaggtctt tgaagtaatt ctgaactccc ctgtgaatgc 540
agttcttggc acaaagacaa aagctgcagt gaaaattttg gactcaaaag gaggacaatg 600
ccatccttca tattcctcca accaaagcaa gcacagcaca tgggagaagg gcatttggca 660
tctgctgccc ccagggtctt cctcatccac cacttctggt tcctttcatc tggaaagaag 720
acctcttcca tcttccatgc agctagcagt catcagggga gacaccctgc ggggctttga 780
ttctacagat ctttctcaaa ggaagcttag gacccgtggg aatggcaaaa cagttcgtcc 840
atcctctgtt tatagaaatg gaacagacat catctataat tatcatggga tagtttcctt 900
gaaactggag gatgacagtt tcccaactca caaaaggaag gccaaagtat ccatcattag 960
tcagccacaa aagacaatca aagtggcaga actgcctcaa gcagataagg tggaatccac 1020
aactgactca cacttcccca gacaggacca gttgccctca tttccaaaga actgcactct 1080
ggaattaaag ggactcttcc attttgaaga aggcatccag aagctgtatc agtgcaatgg 1140
gatcgcctgg aaagcctgga gtccccaaac caaggatgtg gaagacaaat cctgtccagc 1200
cgggtggcac cagcactcag gctactgtca catcttgatc acagagcaga aaggcacctg 1260
gaatgcggct gcccaagctt gcagggaaca atacctgggc aaccttgtaa ctgtattctc 1320
caggcagcac atgcggtggc tctgggacat tggtgggaga aagtcctttt ggataggttt 1380
gaacgaccaa gtgcatgctg gccactggga gtggatcggt ggtgaacctg ttgccttcac 1440
caatgggaga agagggccct ctccacgctc caagcttgga aagagctgtg ttttggttca 1500
aagacaaggg aaatggcaaa caaaagactg taggagagcc aaacctcata attatgtgtg 1560
ttccagaaaa ctctaaatat aacagaccct acaggggccc acctggagtt tgtcacctat 1620
ttattcacag gatctgtgaa tattgctcca tagaaaacaa attgttatga ttgagtgggt 1680
atacctttgt gattctgtct agtgaaaatg ggacattttt aatagtgcca gaaagattga 1740
taaataaata ttttttacaa gataagatac aatttttgta tctcaatacc ttttaaaata 1800
aatgccagca gtattaaaaa aaaaa 1825

<210> 50
<211> 1712
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4675668CB1

<400> 50
```
ctttcttcag tccccacgtg cgatccttcc cggcaacttt ttcgagaaaa atgcccaaat   60
tcaaggcggc ccgtggggtg gggggtcagg aaaaacatgc gcccctggcc gatcagatcc  120
tggctgggaa tgcggtgcgg gcgggggtcc gggagaagcg gcggggtcgc gggacaggag  180
aagcggagga agagtatgtg gggcccggc tgagccgacg gattttgcag caagcacggc  240
agcaacagga ggaactcgag gccgagcatg ggactgggga caagcccgcg gcgccgcggg  300
aacgcaccac gcggctgggt ccaagaatgc ctcaggatgg atcagatgac gaggacgagg  360
agtggcccac cctggagaag gctgccacaa tgacagcagc gggccatcat gcagaggtgg  420
ttgtggaccc tgaggatgag cgtgccatag agatgttcat gaacaagaac cctcctgcca  480
ggcgcaccct ggctgacatc atcatggaga agctgactga gaagcagaca gaggttgaga  540
cagtcatgtc agaggtgtcg ggcttcccta tgccccagct ggaccccggg gtcctagaag  600
tgtacagggg ggtccgggag gtattatcta agtaccgcag tggaaaactg cccaaggcat  660
ttaagatcat ccctgcactc tccaactggg agcaaatcct ctacgtcaca gagccggagg  720
cctggactgc agctgccatg taccaggcca ccaggatttt tgcctctaac ctgaaggaac  780
gcatggccca gcgcttctac aaccttgtcc tgctccctcg agtacgagat gacgttgctg  840
aatacaaacg actcaacttc catctctaca tggctctcaa gaaggccctt ttcaaacctg  900
gagcctggtt caaagggatc ctgattccac tgtgcgagtc tggcacttgt accctccggg  960
aagccatcat tgtgggtagc atcatcacca agtgctccat ccctgtgttg cactccagtg 1020
cggccatgct gaaaattgct gagatggaat acagcggtgc caacagcatc ttcctgcgac 1080
tgctgctgga taagaagtat gcactgcctt accgggtgct ggatgcccta gtcttccact 1140
tcctggggtt ccggacagag aagcgtgaac tgcctgtgct gtggcaccag tgcctcctga 1200
ctttggtcca gcgctacaag gccgacttgg ccacagacca gaaagaggcc ctcttagaac 1260
tgctccggct gcagccccat ccacagctat cgcccgaaat caggcgtgag cttcagagtg 1320
cagtcccccg cgatgtggaa gatgttccca tcaccgtgga gtgaggaaaa cagtcagctg 1380
tcctggccaa aggggtttgg aaggacacca agacccccgt tggtgactga agatgacact 1440
gagctttaat ggctgaagac ccagatcagg gcagtgacag atcacaggga catctgtggc 1500
tcccagtcca ggacaggaag gactgagggt ctggctggtt ccctcttcca ttctaggccc 1560
ttatccctgt ttagttctga gagccaactt gagataccat atgctagcat tcccagtccc 1620
cagctggggc ttggtgtgag tactttttct atggctattg tgtcaggtca ctgtggataa 1680
aggcaaagac agatatttat tgaaaaaaaa aa                               1712
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 November 2000 (16.11.2000)

PCT

(10) International Publication Number
WO 00/68380 A3

(51) International Patent Classification[7]: C12N 15/12, C07K 14/78, 14/47, C12N 15/63, A01K 67/027, C07K 16/18, C12Q 1/68, A61K 38/17, G01N 33/68

(21) International Application Number: PCT/US00/12811

(22) International Filing Date: 10 May 2000 (10.05.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/133,643	11 May 1999 (11.05.1999)	US
60/150,409	23 August 1999 (23.08.1999)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/133,643 (CIP)
Filed on	11 May 1999 (11.05.1999)
US	60/150,409 (CIP)
Filed on	23 August 1999 (23.08.1999)

(71) Applicant *(for all designated States except US)*: INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #12, Mountain View, CA 94040 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94086 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

(88) Date of publication of the international search report: 19 April 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/68380 A3

(54) Title: EXTRACELLULAR MATRIX AND ADHESION-ASSOCIATED PROTEINS

(57) Abstract: The invention provides human extracellular matrix and adhesion-associated proteins (EXMAD) and polynucleotides which identify and encode EXMAD. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of EXMAD.

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/78 C07K14/47 C12N15/63 A01K67/027
C07K16/18 C12Q1/68 A61K38/17 G01N33/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K C12N A01K C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, EPO-Internal, WPI Data, BIOSIS, CHEM ABS Data, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] Accession number AI188216, 14 October 1998 (1998-10-14) ROBERT STRAUSBERG: "qd66g12.x1 Soares_testis_NHT Homo sapiens cDNA clone" XP002146658 the whole document ---	3,5-8, 10-14
A	WO 99 00410 A (INCYTE PHARMACEUTICALS, INC.) 7 January 1999 (1999-01-07) the whole document ---	1-17,20, 23

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 September 2000	22.12.2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	MONTERO LOPEZ B.

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE EMBL [Online] Accession number AF151838, 1 June 1999 (1999-06-01) XP002146659 the whole document & LAI C.-H. ET AL.: "Identification of novel human genes evolutionarily conserved in Caenorhabditis elegans by comparative proteomics" GENOME RESEARCH, vol. 10, no. 5, May 2000 (2000-05), pages 703-713, -----	1-17,20, 23

INTERNATIONAL SEARCH REPORT

ational application No.
PCT/US 00/12811

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claim 16 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the composition.

2. [X] Claims Nos.: 18, 19, 21, 22
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-23 (partially)

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 18, 19, 21, 22

Present claims 18, 19, 21 and 22, directed to agonists and antagonists relate to an extremely large number of possible compounds. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is not to be found, however, for any specific example of the compounds claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, no search has been carried out for claims 18, 19, 21 and 22.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-23 partially

   Polypeptide comprising SEQ ID NO:1, variants and fragments thereof, antibody binding to it; polynucleotide of SEQ ID NO:26, variants thereof, cell and transgenic organism comprising the same; probes derived from the polynucleotide and use thereof in a diagnostic method; pharmaceutical composition comprising the polypeptide and its therapeutic use; use of the polypeptide in screening assays for agonists, antagonists and compounds capable of altering the expression of the polynucleotide; therapeutic use of the agonists and antagonists.

2. Claims: 1-23 partially

   Idem as subject 1 for SEQ ID NOs:2 and 27

3. Claims: 1-23 partially

   Idem as subject 1 for SEQ ID NOs:3 and 28

4. Claims: 1-23 partially

   Idem as subject 1 for SEQ ID NOs:4 and 29

5. Claims: 1-23 partially

   Idem as subject 1 for SEQ ID NOs:5 and 30

6. Claims: 1-23 partially

   Idem as subject 1 for SEQ ID NOs:6 and 31

7. Claims: 1-23 partially

   Idem as subject 1 for SEQ ID NOs:7 and 32

8. Claims: 1-23 partially

   Idem as subject 1 for SEQ ID NOs:8 and 33

9. Claims: 1-23 partially

   Idem as subject 1 for SEQ ID NOs:9 and 34

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

10. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:10 and 35

11. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:11 and 36

12. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:12 and 37

13. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:13 and 38

14. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:14 and 39

15. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:15 and 40

16. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:16 and 41

17. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:17 and 42

18. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:18 and 43

19. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:19 and 44

20. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:20 and 45

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

21. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:21 and 46

22. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:22 and 47

23. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:23 and 48

24. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:24 and 49

25. Claims: 1-23 partially

    Idem as subject 1 for SEQ ID NOs:25 and 50

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9900410 A	07-01-1999	US 5872234 A	16-02-1999
		AU 8160898 A	19-01-1999
		EP 0988318 A	29-03-2000

THIS PAGE BLANK (USPTO)